FUNDAMENTAL FREQUENCY DISTRIBUTIONS OF BILINGUAL SPEAKERS IN FORENSIC SPEAKER COMPARISON

A dissertation submitted

to The University of Canterbury in partial

fulfillment of the requirements for the

degree of Master of Linguistics

by

Kieran Dorreen

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all I would like to thank Vica Papp, without her constant support and motivation there is no way I would have completed this thesis.

I would also like to thank Jeanette King and everyone else involve with the MAONZE project for allowing me to use their corpus in this study. Thanks also to Liam Walsh for collating the QuakeBox recordings for me, and to Robert Fromont for applying REAPER on to all of the recordings.

Lastly I would like to thank Jesse Sheehan, without him and his God-like computer skills I would probably still be without usable data.

CHAPTER 1

Introduction

The fundamental frequency (f0) is the main acoustic correlate of the voicing pitch produced by a speaker. Since the beginning of forensic speaker comparison research f0 has been a popular measure, however its highly variable nature has led to many years of debate on its usefulness as a speaker discriminant. Nolan (1983) suggests six parameters to be used in speaker comparison:

1. High between-speaker variability
2. Low within-speaker variability
3. Resistance to attempted disguise or mimicry
4. Availability
5. Robustness in transmission
6. Measurability

Nolan argues that f0 is an accurate speaker discriminant, asserting that it complies with the above requirements. He lauds especially its availability, measurability, and its robustness. However, there is a large debate on f0's effectiveness when it comes to the first two parameters: between-speaker variability and within-speaker variability. F0 may be a common, easily measurable discriminant, but it is also shown to be highly variable within-speaker. This alone would make it useless as a speaker discriminant as two different voice samples of the same speaker could potentially have two very different f0 measurements. Despite this f0 is still one of the most popular speaker discriminants used by forensic phoneticians (Gold 2014).

Forensic phoneticians have until recently largely ignored what may be a substantial portion of the fundamental frequency range (see section 2.3). Creak phonation sits in a speaker's lower frequency range, a range that is often too low for pitch tracking tools to accurately discover. However, with new advancements in the accuracy of pitch tracking it may now be possible to accurately measure creak within a speaker. This has major implications for the analysis of f0 as an entire phonation type that was essentially ignored and / or mismeasured can now be systematically evaluated.

In this thesis I will present a new way to quantify the fundamental frequency within forensic speaker comparison. In chapter 2 I will outline the current state of fundamental frequency analysis in the field of forensic speaker comparison. F0 analysis has been marred by inconsistent practices and inaccurate data collection methods stemming from the use of imprecise pitch tracking. To rectify these methodological inconsistencies I will propose a new method of f0 analysis in chapter 3. This new method starts with a new, more robust pitch tracker that can accurately track the lower frequency ranges. This new method see creak phonation data being analyzed alongside the modal f0 data in a way that was impossible before without accurate pitch tracking. I will test this new method in chapter 4, looking at the within-speaker variation and the between-speaker variation of two sets of bilingual speakers: a homogenous group of bilingual speakers and a heterogeneous group of bilingual speakers. The recordings of these bilingual speakers are ideal for examining f0 variation both within-speaker and between-speaker as all conditions except the language they speak in remain the same. Chapter 5 will present small case studies of three speakers who had substantially different frequency distributions than the rest of the speakers in the two corpuses. I will discuss the implications these outliers may have on f0 analysis and forensic speaker recognition. And finally in chapter 6 I will summarize the results of the proposed method of f0 analysis and suggest a way forward for f0 analysis in forensic speaker comparison.

CHAPTER 2

Literature Review

The use of pitch in forensic speech science has been discussed and debated since the formation of the field. The general consensus is that while pitch is a robust measurement, the variable nature of it means that it is only marginally reliable in practice. However, most studies that discuss the nature of pitch are very limited in their analysis and often disregard pitch features such as creak. Recent studies have explored the potential of creak within new parameters and discuss the potential more detailed pitch analysis has on forensic speech science. This section will outline the history of pitch analysis in forensic speech science and discuss the modern advances in both pitch measurement and analysis.

2.1 Quantifying f0

The fundamental frequency (f0) is the frequency of vibration in the vocal cords in phonated speech, and is perceived as pitch (Rose, 2002). F0 has many qualities that make it an attractive option for speech comparison, such as its robustness, measurability, and its availability within a speech sample (Nolan, 1983). Speakers show distinct differences in their distribution of spectral energy in their speech, largely due to anatomical differences and how individual speakers manage their vocal tract settings (Clark and Yellop, 2001). It is for this reason that f0 analysis is a common measurement extracted in speaker comparison. However, despite these attractive features there are many more that detract from the overall effectiveness of pitch analysis in speaker comparison. External factors such as emotion (Maekawa, 1998, Paeschke et al., 1999), health (both short term illness (Rose, 2002) and long term illness (Bowen et al, 2013)), disguise (Künzel, 2000), and volume (Elliot, 2000) have been shown to have a significant effect on a speaker's f0, showing that f0 is a very variable feature within the speaker. For example Elliot (2000) showed that the mean f0 and standard deviation around the mean (SD) were higher in shouted speech than in normal speech. Maekawa (1998) showed that the mean f0 and SD were variable within Japanese speakers' realizations of paralinguistic information, e.g. admiration, disappointment, or indifference. While both these studies found within-speaker f0 variability, these studies (like most of the studies mentioned above) suffer from small sample sizes (2 and 3 respectively) and they both were completed under laboratory conditions, which is not ideal for forensic speech research as most data relating to that is composed of natural speech.

With the exception of Rose (2002), who combined other parameters such as the mode, skew, and kurtosis, all studies presented thus far have only used two methods of measurements: the mean f0 and the standard deviation (SD). This appears to be the standard throughout the f0 literature, be it related to forensic, clinical, or linguistic applications. Until recently it appeared that mean f0 and SD are the only ways f0 was quantified in forensic contexts. In most studies the mean f0 and SD showed highly variable results for within-speaker variation, which drives its reputation of a week speaker discriminant. A case study by Boss (1996) exemplifies the general conclusion among forensic phoneticians that the f0 vocal feature is too variable for real world application. Boss showed that the use of mean f0 was ineffective in comparing two speech samples of a real life criminal. A recording of a gas station hold up had the mean f0 of the criminal at 228Hz, while an audio recording made after his arrest had his mean f0 significantly lower at 140Hz. Boss warns against the use of f0 within the field as results can be far more variable than those seen in controlled experiments. Studies repeatedly demonstrated that if the analysis stops at the mean f0, the use of pitch as a forensic tool can be disregarded as ineffective. It is possible that the use of a more detailed analysis with a greater number of f0 parameters could have affected the outcome of this case study, and other studies involving the use of f0 as a speaker identifier. However, despite the general bad reputation f0 analysis has within forensic speech science, Gold's (2014) survey of expert forensic phoneticians found that all 36 experts use or have used f0 analysis in their work, with a general consensus that f0 analysis was one of the most useful speaker discriminant alongside voice quality.

2.2 Between-speaker Variation

While within-speaker variation is an important part of the forensic speaker comparison literature, it is more common to see studies evaluating f0 in forensic speech science by analyzing the spontaneous speech of a large homogenous group of speakers (Hudson, 2007). This generalization of speaker groups allows for the comparison between a single speaker and a larger group or population, but avoids the issues discussed above on the variable nature of f0 within a speaker. Fundamental frequency statistics for both female and male speakers across multiple languages can be seen in Traunmüller and

Eriksson (1995). This study provides evidence that f0 is language specific by showing differences in the mean f0 of a population of speakers instead of just one.

Gold (2014) draws comparison between the f0 studies conducted by Rose (2002) and Loakes (2006). Both of these studies provide measurements of the f0 of Australian men. Rose (2002) recorded the read speech of six Australian men on two different occasions and showed that their mean f0 and standard deviations were 113.6Hz and 21.7Hz on the first recording, and 114.5 Hz and 17.4Hz for the second recording. Loakes (2006) measured the f0 at the midpoint of the vowels in eight-minute recordings of spontaneous speech. He found the mean f0 of Australian men to be lower in his study, at 105.2Hz (SD 16.4 Hz). Loakes attributes this to the fact that his speakers provided spontaneous speech, while the speakers in Rose (2002) were measured with read speech.

Both these studies exemplify the issues that surround this type of analysis: there has been a larger focus on between-speaker f0 variation in lieu of work done on within-speaker variation. This is possibly due to the variable nature of pitch in a speaker. Instead pitch is treated as a single value representing a group of speakers, disregarding the variable nature of pitch between and also within the speakers. Gold (2014) illustrates this with two studies of large speaker groups. Firstly Lindh's (2006) study of 109 young male Swedish speakers reports their mean f0 as 120.8Hz and a median f0 of 115.8Hz. Hudson et al. (2007) look at 100 recordings of male British English speakers simulating police interviews and reports a mean F0 of 102.2Hz, a median F0 of 106Hz, and a mode F0 of 105Hz. Both these studies attempted to gain understanding of f0 distributions within a large, homogenous group of people. The comparison of Lindh's Swedish speaker study with the other English speaker studies provides further evidence that f0 is language specific.

This is not to say that within-speaker variation has been neglected as a whole in fundamental frequency studies. While between-speaker variation is clearly a more popular analysis, within-speaker variation has been increasingly gaining traction over the past few years, particularly in conjunction with between-speaker variation. Kinoshita (2005) was one of the first researchers to co-evaluate the two, as both between- and within-speaker variation are needed for the numerical likelihood ratio based framework that was being applied in this study. Calculating likelihood ratios shows the strength of forensic evidence; instead of just showing that two samples are different or the same, this methodology of expressing conclusions applies a likelihood ratio (LR) test to determine how different or similar the two samples are. Kinoshita's study of 90 Japanese men reported a mean f0 of 135.7Hz and a standard deviation of 26.4Hz. Separate 'criminal' and 'suspect' f0s were created using the original 90 speakers as a reference population. Likelihood ratios were calculated but the results were inconclusive, because the LRs were unable to determine if a speaker in the 'criminal' recordings were the same in the 'suspect' recordings. Kinoshita concludes that f0 is not a strong speaker discriminant as it shows little strength of evidence.

In a study on bilingual speakers Voigt et al. (2016) showed the effects of both between and within-speaker variation. There were 45 speakers in total: 22 men and 23 women. 25 were German-French bilinguals while the remaining 20 were German-Italian. Voigt et al. showed that there is variance between the two language groups in their f0 readings when they are speaking German, and within the speaker there is variation in their f0 depending on what language they are speaking. This observation was found to be consistent within all speakers. A gender effect was also observed, with the amount of f0 variance appearing to be determined by the gender of the speaker and what language they were speaking. This implies that f0 variance is consistent across different languages, and that this effect may be partly due to gender as well as language. This study only used the mean f0 in the analyses, however, it did not provide pitch values to the specific speakers or language groups.

2.3 Issues with Creaky Phonation

Creaky phonation occurs when the vocal folds are tightly compressed yet open enough to allow for voicing (Gordon and Ladefoged, 2001). This makes for a slower speed of glottal cycles, which puts creaky phonation on the lower end of an f0 distribution. It is well documented that creaky phonation is a normal part of speech in most speakers, regardless of gender, language, or other factors (Abdelli-Beruh et al., 2013; Wolk et al., 2012; Yu and Lam, 2014; Aare et al., 2014). However, most studies disregard creak phonation in their analysis of f0. Both Hudson (2007) and Gold (2014) note that in most cases an f0 distribution is clearly bimodal: there are two distinct peaks of frequent pitch values, one for creaky phonation and the other for modal phonation. However, after detailing the bimodal nature of the f0 distribution both these studies proceeded to disregard creak as a separate measurement and combined both creaky phonation and modal phonation ranges into a single mean f0 calculation. Gold (2014) does suggest finding a way to incorporate bimodality into f0 analysis in forensic speaker comparison. Kinoshita and Ishihara (2012) tested different parameterization techniques to capture effects of non-unimodal distributions (f0 distributions with more than one peak). They found that taking into account non-unimodal distributions had significant positive effects on the produced likelihood ratios.

There is no standard methodology when dealing with creak in an f0 distribution. As seen above, some studies keep creak in the calculations and treat the entire distributions as a single object (Hudson, 2007; Gold, 2014). There have been attempts to include creak as a separate feature in the overall analyses (Kinoshita and Ishihara, 2010). Some omit creak altogether and only focus on modal phonation (Leeman et al., 2014). However, the majority of studies (including mostly all studies previously mentioned) don't recon with creak at all, making it unknown how creak is considered in the study. The omission of such information gives ambiguity to measurements such as the mean f0 as the amount of creak has the potential to shift this value based on the range quantified. The mean f0 would be lower in a speaker whose creak was included in the mean f0 calculation than if it was omitted.

The justification for excluding creaky phonation data is that pitch tracking is unreliable in the lower pitch ranges. Gold (2014) argues that since pitch tracking in the lower pitch ranges is so unreliable the effects of the few creak measurements that will be picked up are negligible. Leeman et al. (2014) however states that the lower pitch tracking is so ineffective in the lower pitch ranges that it portrays an inaccurate picture of creak so therefore should be excluded. Both these studies made use of Praat (Boersma and Weenick, 2013) for pitch tracking, by far the most popular pitch tracker used in f0 studies.

As ineffectual as it is, Praat is the current standard for pitch tracking. This is problematic, as it does not provide accurate data, especially in the lower frequency ranges. However, a new pitch tracker, named REAPER (Robust Epoch and Pitch EstimatoR (Talkin, 2015)) shows promise as an effective and very accurate pitch tracker. It works by estimating the location of voiced speech “epochs” (glottal closure instants), voicing state, and the fundamental frequency. Liberman (2015) tested REAPER and showed that it was both accurate and effective in tracking the pitch of a notoriously creaky voiced radio personality, Ira Glass. REAPER accurately tracked the pitch down to 27Hz, a reading which would be virtually impossible to attain in Praat.

Overall creak has been generally ignored in most f0 studies, mostly due to inaccurate pitch tracking methods. However, as Kinoshita and Ishihara (2010) shows, the inclusion of creak in f0 analyses can have a significant positive effect on measures such as likelihood ratios. Advancements in pitch tracking also allow for more accurate f0 analyses.

2.4 Case Study: Gold (2014)

Gold’s PhD thesis (2014) summarizes the field of forensic phonetics, exposing in detail the methodological inaccuracies and inconsistencies prevalent, specifically concerning fundamental frequency analyses. Gold polled a group of 36 forensic phoneticians about their ideas and techniques in the field. Gold’s own laboratory study used 100 male British English speakers with recordings ranging from 2:25 minutes to 11:17 minutes long (with an average length of 6:21 minutes). Each speaker's recording was split into 10-second segments (in order to calculate likelihood ratios, which need multiple tokens per speaker) and the mean f0 and SD were measured for each token using Praat.

The first issue with this methodology is Praat's handling of the pitch tracking in order to calculate the mean f0 and standard deviation for each token. The Praat script used in this study was set to a frequency range of 50-300Hz. A review of the initial Praat outputs found that 64 speakers had what was considered "reliable" f0 measurements, the remaining 38 contained erroneous readings such as octave jumps and had to be redone using pitch ranges tailored through trial and error for each speaker. Gold reports that the range with least amount of errors was used for each speaker, indicating that there were still errors present even after a large amount of manipulation of a section of the data.

In an effort to standardize the large group of speakers Gold uses multiple levels of calculations, the most abstract of which is the "mean of the mean of the mean." This is defined by Gold as "the mean across speakers of the means across tokens of each speaker of the means of all the raw F0 values of each token" (Gold 2014, pp. 190). This type of calculation not only potentially obscures cases of individual variation within a speaker but it can also be affected by the amount of creak a speaker uses, an issue is not mentioned by Gold in the discussion of her methodology. The entirety of the f0 distribution was used in the calculation of each speaker's f0.

Gold admits that the mean f0 is language specific. Indeed her data shows that all 100 speakers are relatively close in range. However, she also indicates that the pitch of her speakers may have been affected by the amount of creak each speaker used. Creak in this study is creak measured by Praat, which is already an inaccurate representation of how much creak a speaker uses. However, because creak phonation and modal phonation were conjoined in the initial measurement phase, there is no way to know how much creak affected each speaker's production, if at all. Gold does suggest the possibility that only some speakers use creak and that it is affected by within-speaker creak variation. If this were true it would potentially magnify inaccuracies in the data even more. Again, there is no way to tell this based on the way the data and the research methods are presented.

Gold acknowledges the shortcomings of her methodology, and those of forensic phoneticians in general. She accepts that the use of creak in her study is not ideal, both in the flawed data collection methods and in how creak is represented in the total f0 distribution. She suggests finding a way to represent bimodality in f0 distributions as the way forward in f0 studies, both forensically motivated and otherwise.

2.5 F0 in Bilingual Speakers

The general consensus in fundamental frequency studies is that f0 is language specific (Rose, 2002; Hudson, 2007; Gold, 2014), and there have been many studies conducted on the f0 of different language speakers. Traunmüller and Eriksson (1995) being the most comprehensive example, but there are many more that focus on specific languages and comparing two languages together (e.g. Connell, 2002; Ackerman et al., 2011; Pepiot, 2014; Luo et al., 2016). What is less common are studies on the effect of f0 on different languages spoken by the same speaker.

Järvinen et al. (2013) investigated whether using a foreign language had an effect on the speaker's f0. 30 speakers were used in this study, 16 native Finnish speakers and 14 native English speakers. The mean f0 and standard deviation were recorded for all speakers in both languages. It was found that the Finnish speakers showed a change in mean f0 when speaking English, but the result was not reciprocated with the English speakers. It was speculated that the shift in mean f0 was the result of adapting to a foreign environment. There was no significant correlation with a speakers experience in using the foreign language, indicating that the mean f0 shift was on the individual level.

Schwab and Goldman (2016) compared English-French, English-German, and French-German bilinguals and applied a within-speaker design to the long-term f0 distribution to determine speaker variation across languages. The mean f0 was obtained through Praat for each speaker along with other f0 parameters such as skew and kurtosis. They found that both English-French and English-German speakers had a lower f0 when speaking English, with English showing more variation especially when English was the speaker's native language. French-German speakers showed no differences in f0 between both languages. These results highlight the usefulness of a within-speaker analysis

Voigt et al. (2016) showed the effects of both between and within-speaker variation. There were 45 speakers in total, 22 being men and 23 women. 25 were German-French bilinguals while the remaining 20 were German-Italian. Voigt et al. showed that there is variance between the two language groups in their f0 readings when they are speaking German, and within the speaker there is variation in their f0 depending on what language they are speaking. This observation was found to be consistent within all speakers. A gender effect was also observed, with the amount of f0 variance appearing to be determined by the gender of the speaker and what language they were speaking. This study used the mean f0 again in its analysis, however it did not provide pitch values to the specific speakers or language groups.

Overall these studies show that in general a bilingual speaker will show some variance in their f0 depending on the language they are speaking. This is not always the case however, as some examples such as the French-German speakers from Schwab and Goldman (2016) showed no significant difference in their f0 measurements. However, most these studies relied on Praat as their pitch tracker, and did not explicitly document the decisions about the inclusion of creaky phonation in the measurements. Since creak has been argued to be language specific it is possible that the amount of creak in a speakers two languages are affecting (masking or amplifying) the overall comparison between the two languages.

2.6 New Approaches to Quantifying f0

The vast majority of studies discussed above use simply the mean f0 and standard deviation as a speaker discriminant, to limited and varying success. However, there is more to the characterization of an f0 distribution than the mean f0 and its SD. Rose (2002) describes other potential f0 discriminants such as the mode (the most frequent f0 value in a frequency distribution), the skew (how wide / tailed the distribution is) and the kurtosis (the height / peakiness of a distribution). These parameters are understudied, as the large majority of f0 studies stop at calculating the mean f0 and standard deviation, regardless of how inconclusive those parameters can be. Kinoshita and Ishihara (2014) used the features discussed in Rose (2002) alongside the mean f0 with some degree of success. Their study looked at the background population and its effect on likelihood ratios in forensic voice comparison and uses multiple parameters (such as mode / skew / kurtosis) to calculate the likelihood ratios. Using multiple parameters showed significant improvements on speaker comparison against background populations, predictably with accuracy ratings higher than when only mean and SD were used. However, creak was not discussed in this study either, so it is unknown how successfully the authors tracked creak and if it was kept in the total frequency distribution or if it was discarded.

Through all the methods discussed and all the potential parameters available there appears to be one consistent struggle: how to account for creak phonation. This issue is expanded with greater accessibility to more accurate pitch trackers such as REAPER, as the more creak that is included in an f0 distribution the more likely it will interfere with traditional f0 mean measurements. One potential way to deal with creak phonation is to treat it separately from modal phonation. Gold (2014) postulates that the amount of creak is individual to the speaker; this may be useful in speaker comparison. There may also be patterns and correlations between creak phonation and modal phonation, which till this point have been hidden or obscured due to either lackluster pitch tracking in the lower frequency ranges or being combined with modal phonation altogether. Therefore I propose to split each speaker's frequency distribution into two – one creak phonation distribution and one modal phonation distribution.

To test this the following study will use recordings of bilingual speakers processed with the REAPER pitch tracker.

2.7 Predictions

Based on the current literature on f0, several predictions can be made on the proposed methodologies and the role of creak phonation in bilingual speakers.

Prediction 1: The use of the REAPER pitch tracker will show significant improvements on the accuracy of current pitch trackers.

REAPER has not been used in a comprehensive study of f0, however in an informal small-scale study Liberman (2015) shows it to be extremely effective at estimating pitch measurements especially at lower pitch ranges. The most widely used pitch tracker (Praat) is well known for inaccuracies in pitch estimation in the lower ranges, and many studies are affected because of it as the inaccurate pitch measurements have the potential to affect the overall results.

Prediction 2: There will be significant variation in f0 between an individual speaker's two languages.

The general consensus among f0 studies is that f0 is language specific, however most studies comparing different language f0 focuses on groups of monolingual speakers, providing a single mean f0 and standard deviation for each group. These means do show significant differences in the mean f0 values for each language group. Recent studies of the effect of bilingualism on a speaker's f0 do show significant differences between a speaker's two languages, though the amount of difference does appear to be constrained to the specific language pair. These studies are limited to both inaccurate pitch tracking and restricted analysis, only using mean f0 and standard deviation. Regardless of this there is little reason to believe that variation between languages will be different in the current study.

Prediction 3: A homogenous group of bilingual speakers will show less variation in f0 than a heterogeneous group of bilingual speakers.

This prediction carries on from prediction 2, which is based on the idea that f0 is language specific. Two corpora will be used in this study, one of a homogenous group of bilingual speakers (speakers that all speak the same two official languages of New Zealand) and another of a heterogeneous group of bilingual speakers (speakers that each speak two different languages). If f0 were language specific it would be expected that the results of the homogenous group are relatively consistent across all speakers, however the heterogeneous group would not be as the speakers demonstrate a variety of different languages.

Prediction 4: Creak phonation alone will be less useful than modal phonation as a speaker discriminant, however its relationship to modal phonation may prove useful.

Creak phonation falls into a smaller range than modal phonation and is considered to play a smaller role in a speakers total f0 distribution (Rose, 2002; Gold 2014). There is more room for variation in all f0 parameters (including skew and kurtosis) in modal phonation so it can be considered a better candidate for speaker comparison. This isn't to say creak will be useless, rather less useful than modal phonation. There are potential relationships between creak and modal phonation that so far have been unable to be explored due to inaccurate pitch tracking and methodologies that connect both creak and modal phonation into a single f0 measurement.

Prediction 5: The addition of further f0 distribution parameters (mode/skew/kurtosis) will improve the accuracy in quantifying the variability of both in the within-speaker and between-speaker condition for use in speaker discrimination.

Nearly all studies concerned with f0 distribution are limited by making use of only the mean f0 and standard deviation parameters. These parameters are considered to only be mildly effective in speaker comparison and discrimination studies as it is highly variable within a speaker. Rose (2002) identified mode, skew and kurtosis that could be potentially fruitfully used for f0 analysis. Kinoshita et al. (2005) showed that when using these parameters together there were significant improvements in the accuracy of likelihood ratios. By using more than just one parameter it is expected improve f0's speaker discrimination powers.

CHAPTER 3

Methodology

Two corpuses of bilingual speakers were used in this study. The first corpus was the Maori and New Zealand English (MAONZE) corpus. This body of recordings contains the speech of bilingual speakers of Maori and English. The second corpus, the QuakeBox Corpus, is not limited to one particular pair of languages. Instead it has an assortment of second languages, ranging from European languages to East Asian languages talking about their earthquake experience in 2010-11. Both corpora are processed and stored in LaBB-CAT, a browser based linguistics research tool that holds audio files, transcripts and other annotations (Fromont and Hay, 2008).

3.1 The MAONZE Corpus

The first half of the speakers in this study come from the Maori and New Zealand English (MAONZE) corpus. This corpus comprises of recordings of three generations of Maori speakers dating back to speakers born in the late 19th century, many of which were recorded speaking both Maori and English. Later recordings in this corpus were purposefully recorded in both languages (King et al. 2010). All usable recordings from speakers born after 1969 were used in this study, resulting in a total speaker count of 17. Table 1 details the speakers from the MAONZE corpus.

Table 1 - MAONZE speakers and their total f0 measurement count

MAONZE Speakers					
Male			Female		
Speaker	Language	f0 Count	Speaker	Language	f0 Count
M01	English	165052	M10	English	322412
	Maori	182886		Maori	122460
M02	English	90520	M11	English	232836
	Maori	103909		Maori	41734
M03	English	134176	M12	English	88711
	Maori	39076		Maori	184679
M04	English	67418	M13	English	213619
	Maori	231304		Maori	173511
M05	English	26048	M14	English	47908
	Maori	559000		Maori	121146
M06	English	13757	M15	English	49262
	Maori	148280		Maori	249790
M07	English	154685	M16	English	59693
	Maori	140078		Maori	260815
M08	English	111795	M17	English	28552
	Maori	145220		Maori	27134
M09	English	186587			
	Maori	130540			

All speakers have one English language recording and one Maori language recording. This was in most cases done sequentially; they would record the Maori interview and then move on to the English interview. Each interview was up to an hour long for each language and comprised of the interviewee reading from wordlists, reading passages, and being interviewed. The interview was informal in structure, with no set topics discussed. Instead the speaker was usually invited to talk about anything they were passionate about. To assist with the storage and access in LaBB-CAT the recordings were segmented into tracks in of around 3-5 minutes in length. To maintain consistency with all speakers in this study any track with a wordlist or reading passage was omitted. The

Maori reading passages were identified by the lack of pauses and filler words (e.g. um or er) within the transcript.

3.2 The QuakeBox Corpus

The QuakeBox corpus was established in 2012 after the major 2010 and 2011 Canterbury earthquakes (Clark et al, 2016**)**. The corpus contains Cantabrians' recounting their experiences from the earthquakes. Most speakers in this corpus are English monolingual, however when a bilingual speaker entered the booth to recount their story they were asked to repeat it in their other language. These recordings were made sequentially, with a speaker recounting their story in English first and their second language second. This corpus is unique in that it includes speakers telling the exact same story in the exact same setting but in different languages. This allows for direct comparison between the two recordings as only the one variable (the language) has changed. In total 17 bilingual speakers across seven different second languages from the QuakeBox corpus were used in this study. Because the recordings are only single stories all of the recordings are significantly shorter than the MAONZE corpus recordings. Each QuakeBox recording is around 5-15 minutes in length. Table 2 details the QuakeBox speakers and their specific languages.

Table 2 - QuakeBox speakers and their total f0 measurement count

QuakeBox speakers					
Male			Female		
Speaker	Language	f0 count	Speaker	Language	f0 Count
QB01	English	61112	QB10	English	33333
	French	66713		Cantonese	44295
QB02	English	78762	QB11	English	160750
	Japanese	89627		German	164413
QB03	English	180640	QB12	English	27473
	Mandarin	169570		Japanese	39143
QB04	English	92221	QB13	English	30440
	Mandarin	109721		Japanese	55039
QB05	English	25834	QB14	English	116483
	Mandarin	63076		Mandarin	218858
QB06	English	120846	QB15	English	66163
	Maori	200824		Maori	46044
QB07	English	60371	QB16	English	53730
	Maori	142048		Maori	93986
QB08	English	45397	QB17	English	121275
	Punjabi	75481		Maori	183745
QB09	English	35703			
	Russian	49294			

3.3 REAPER

The algorithm REAPER (Robust Epoch And Pitch EstimatoR) was applied to all recordings from both corpuses, a total of 68 recordings. This process was completed on the University of Canterbury servers through LaBB-CAT. As the amount of processing power needed to apply so many large recordings through REAPER was very large. REAPER's output was a text file that contained the f0 value every 0.005 seconds from the whole recording. Voicelessness was recorded as having a value of -1. As the output provided f0 values through the whole recording f0 values of an interviewer or other speaker were also attained. This mostly concerned the MAONZE recordings, which were conversational with the speaker and an interviewer; the QuakeBox recordings were monologues with little to no input from a facilitating interviewer. As the MAONZE recordings were split into tracks on LaBB-CAT the REAPER output for each speaker was spread across multiple result files, which were then concatenated at a later phase. While REAPER is able to track even lower f0 values (see Liberman, 2015), for the purposes of this study the pitch floor was kept at the default 40Hz.

A custom written Python script (shown in full in appendix A) was used to organize the data for both corpuses. The Python script automatically removed every instance of voicelessness in the REAPER output. It also used interval timestamps from the accompanying Praat TextGrids to remove any intervals of an interviewer within the recording. Then the script concatenated all f0 outputs for a single speaker into one file. Lastly, the Python script converted the concatenated REAPER output file to a .csv file for subsequent analysis in R.

3.4 R Plots

All files were then imported into R and probability distributions (density plots) were created for each speaker. An example of a density plot for one speaker (M08) is shown in figure 1. These density plots established that all speakers had a bimodal distribution in at least one language in their speech. In order to account for both distributions it was decided to split them into a separate creak phonation distribution and a modal phonation distribution. To identify the splitting point in the distributions the speaker and language specific first antimode was used. The antimode is the opposite of the mode statistic; it is the point between in distributions that has the least frequent f0 value. In this particular case the antimode between the creak and the modal distribution was established. This was calculated using the specialized *modes* R package (Deevi, 2016).

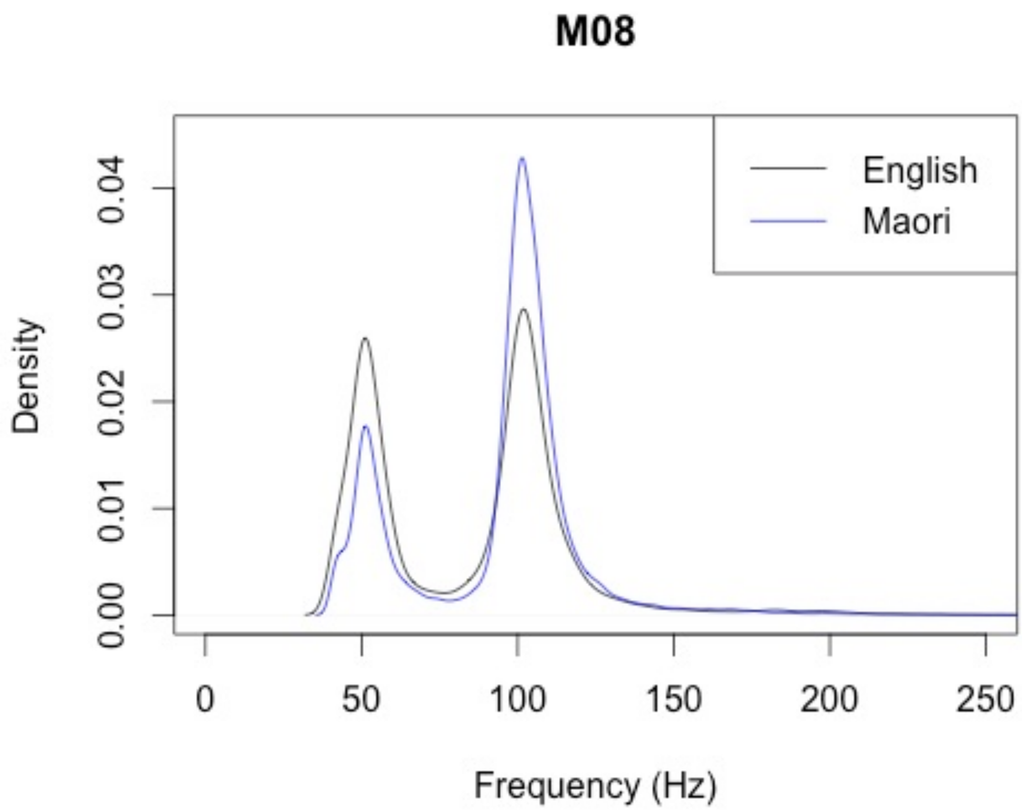


Figure 1 - Frequency distributions of English and Maori for speaker M08

The mean f0 and standard deviation for both creak phonation and modal phonation were calculated for each speaker's two languages in R. The modes, skew, and kurtosis for each phonation type were all calculated using functions from the same *modes* package used to calculate the antimode. The other measurement used in this study is the percentage of creak; this was calculated as the percentage of creak measurements a speaker produced within a recording. While calculating the antimode for splitting the distributions it was found that each speaker shows similar antimode values between their two languages. This warranted a closer look at the antimode as a speaker discriminant, and as such it will be included in the analyses alongside other established parameters such as mean f0 and mode f0.

Bar plots were created using the *ggplot2* package in R (Wickham, 2009) for each parameter showing the parameter values in relation to within-speaker variation and across speaker variation. An example of the R code used in this study can be seen in full in appendix B.

CHAPTER 4

Results

4.1 Modality

The first result from this data shows that no matter the language being spoken there is a clear bimodal distribution in the speakers' pitch. Figures 2 and 3 are examples of density plots produced for all speakers from both corpuses used in this study.

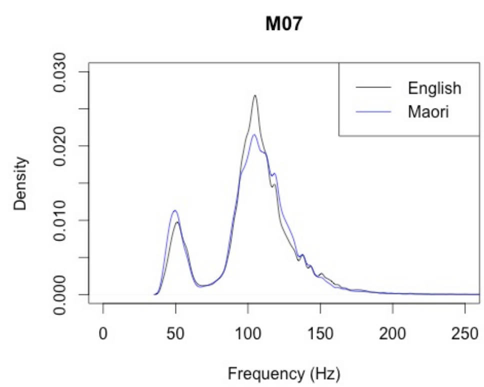


Figure 2 - density plot for M07

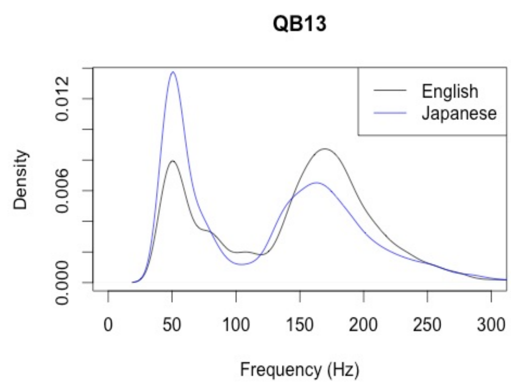
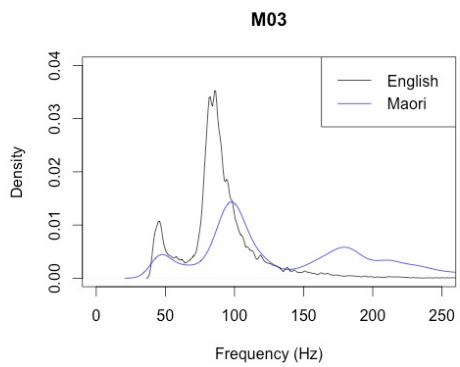


Figure 3 – density plot for QB13

Figure 2 is a Maori-English bilingual male speaker from the MAONZE corpus and figure 3 is a Japanese-English bilingual female speaker from the QuakeBox corpus. The x axis (frequency) shows the fundamental frequency (f0) range for each speaker's two languages, while the y axis (density) displays the number of times (frequency count) a speaker uses a specific pitch value in the 1 Hertz wide bins. Both show the clear bimodal distributions in speech in both their spoken languages. These figures illustrate the general finding in the two corpora that regardless of language or gender the f0 distribution is bimodal. This is apparent in all but two speakers, shown in figures 4 and 5. M03 has a trimodal distribution in his Maori pitch and M06 shows a Maori pitch closer to unimodal than bimodal.



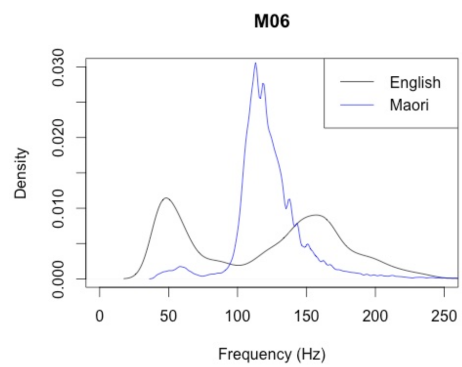


Figure 4 - Density plot for M03

Figure 5 - Density plot for M06

Despite the general bimodality (aside from figures 4 and 6), there is no clear predictable relationship between the f0 probability distributions of the two languages spoken by a speaker. Speaker M07's two languages are very similar in distribution, while QB13 shows that while their English creak and modal distributions are relatively equal, their Japanese shows a much more frequent use of their creak range than their modal range. This lack of relationship is seen in all speakers regardless of gender or language pair. Another noticeable difference between figures 2 and 3 is the smoothness of the distribution. Speaker QB13 shows two clear peaks with a relatively smooth distribution, while speaker M07 isn't so smooth, with various bumps along the right side of the modal distribution that appear in both languages. For the purposes of this study these bumps will be disregarded with each distribution (creak or modal) and only two modes per speaker per language will be considered.

The finding that all speakers, regardless of language, have a bimodal pitch distribution contradicts the research methodology applied in previous studies. Out of the literature surveyed, only Hudson (2007) and Gold (2014) acknowledge that pitch is bimodal yet even they calculate one mean pitch value from a range that includes both the creak and modal distributions. Kinoshita and Ishihara (2012) tested parameterization techniques to account for non-unimodal distributions, but other than these three studies the bimodal nature of pitch distributions appears to be completely ignored. This is concerning as there are a number of statistical data summarization methods (means, standard deviations) that will be incorrect as their basic assumption is a unimodal distribution.

Firstly the often used mean f0 may be significantly lowered when calculated using both creak and modal distributions. This is because the more creak phonation a speaker uses the lower the total mean f0 is going to be. Secondly the mode of f0 measurement can possibly be affected if one single mode is calculated for the entire f0 range used. Since mode of f0 is just the most frequently used pitch value (the highest peaks in the probability distribution plots seen in figures 2 and 3), it is a more robust measurement than mean. This is because the mode of f0 is resilient against outliers and doesn't move around the speakers' distribution as much as the mean f0 does. However, in cases such as figure 3 above, the highest point is in the creak range. If the total mode f0 measurement was taken for the Japanese distribution in figure 3 the result would be somewhere around 50Hz, which is clearly wrong for a female speaker.

Skew and kurtosis, two measures often used to characterize the shape of a unimodal distribution, are erroneously calculated if one disregards bimodality. Both skew and kurtosis measure the shape of a distribution, with skew measuring the width and spread, and kurtosis measuring the height and peakiness. However, there are clearly two distributions within a speakers pitch. In studies where skew and kurtosis has been measured (such as Kinoshita et al. (2009) there is no indication of what the skew or kurtosis is measuring, if it is from the creak phonation, the modal phonation or the complete phonational range. In the statistics literature it is also unclear how to interpret skew and kurtosis in multimodal distributions and if and how they are useful diagnostics at all.

It is clear that disregarding bimodality in pitch analysis can have significant effects on the overall results. In order to account for bimodality this study will present results from the two distributions, creak phonation and modal phonation, separately. The antimode (the least used pitch between the creak mode and the modal mode) of each total distribution will be used as the separation point between the two phonation types. According to this logic all pitch values to the left of the antimode will be considered creak phonation and all pitch values to the right of the antimode will be considered modal phonation.

4.2 Percentage of Creak

The percentage of creak is the number of creaky glottal cycles expressed as a proportion of a speaker's total voiced glottal cycles. This is represented in the bar graphs seen in figures 6 and 7, with figure 6 representing the MAONZE speakers and figure 7 representing the QuakeBox speakers. Each colour on the graph denotes a single speaker, each of which has two bars representing each of the speakers' language. The line connecting the two bars illustrates the distance, or the nearness, of the percentage values seen in each speakers respective languages, and conveniently aids the reader's eyes in matching up the columns.

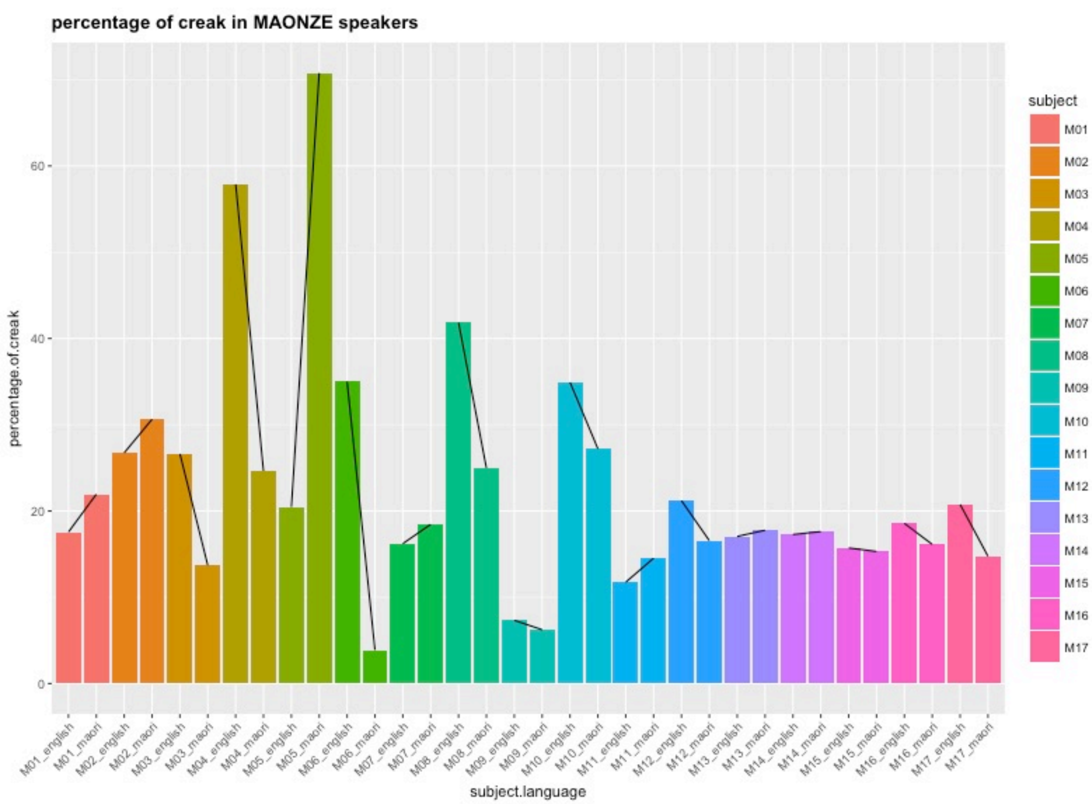


Figure 6 - Percentage of creak in MAONZE speakers

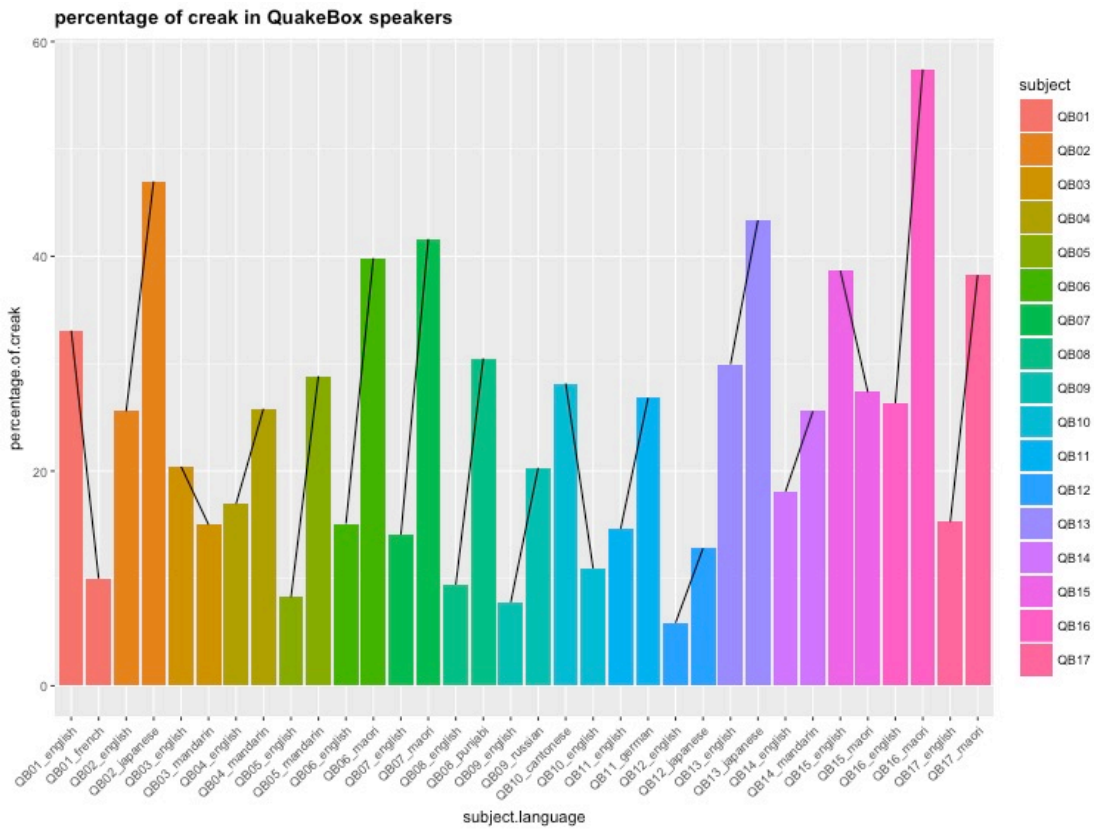


Figure 7 - Percentage of creak in QuakeBox speakers

The MAONZE speakers as a whole seem to use a similar amount of creak in both of their languages. For example speakers M13, M14, and M15 are all near equal when comparing the amount of creak used in their English to their Maori. There are a few speakers who show extensive differences in their creak percentages (namely M04, M05, and M06), but as a whole the MAONZE speakers show strong similarities in the overall amount of creak they use.

The QuakeBox data, however, does not show the same similarities within-speaker creak percentage. Nearly all bilingual QuakeBox speakers show a definite difference between the amount of creak they use in English and the amount they use in their second language. Interestingly, only three speakers produce more creak in English than in their second language, the vast majority of the QuakeBox speakers use creak far more often when speaking in their second language. What accounts for the differences seen in the two corpuses? It is possible that the amount of creak is language (and possibly also dialect) specific. Maori speakers may have the same amount of creak than say, French or Mandarin speakers. This would account for the similar creak percentages for the majority of speakers in the MAONZE corpus and for the variety of differences in the QuakeBox corpus. However, there are five Maori speakers in the QuakeBox corpus, all of which show vastly different creak percentages between their languages. Another possibility is that of recording length. There is one major difference between the MAONZE and the QuakeBox corpuses besides the languages being spoken: the length of the recordings. Each recording for the MAONZE corpus was about an hour, while each QuakeBox recording varied from between 5 to 15 minutes. It’s possible that the more a speaker talks in either language, the more equal the amount of creak phonation becomes between their two languages.

There is no distinct gender difference in the amount of creak produced. Females from the MAONZE corpus show more equal creak percentages (both between their two languages and across speakers) than the male speakers, with the male speakers showing slightly more use of creak. This goes against popular belief of creak, which states that females use more creaky phonation than males (Melvin and Clopper, 2015) The QuakeBox speakers show more variability in their creak percentages than the MAONZE speakers, but as a whole there does not seem to be a major difference between male and female speakers.

Overall the percentage of creak does not appear to be a good discriminant for speaker comparison. Creak is far too variable within the speaker to effectively compare two recordings. However, there are other implications of this large variability, such as its ability to affect other pitch measurements. If, for example, the total mean f0 of two voice recordings of the same person were recorded and the speaker had a higher percentage of creak in the second recording, the second mean f0 would be lower than the first. The amount of creak a speaker uses has the ability to shift the total mean f0. Since the percentage of creak is clearly variable within a speaker this creates the possibility of (rightfully) giving the same speaker two very different total means. This would indicate that the speakers in the two recordings are not the same, when they are the same speaker producing different levels of creak.

4.3 Mean f0

4.3.1 Total Mean

The status quo of quantifying f0 is calculating the mean f0. Generally this entails taking all available pitch measurements between a predefined pitch floor and pitch ceiling (cf. Praat manual) and averaging them to give a single mean f0 value to the speaker in a recording. However, it has been shown that treating the whole of a speaker's f0 distribution instead of taking into account the bimodal nature of pitch can have consequences on the overall mean f0 measurement. What has yet to be determined is by how much the mean f0 can shift within a speaker, and if the difference is significant. This section will look at the total mean f0 from all speakers in both the MAONZE and QuakeBox corpus.

It must be noted that this is the worst-case scenario in calculating the total mean f0. Most pitch tracking studies use software that is not as accurate as REAPER, with Praat's pitch tracking tool being the most widely used. These other pitch tracking tools are especially ineffective within the creak range. This is exemplified in figure 8, which compares the distribution of the same speaker using Praat's pitch tracker and REAPER.

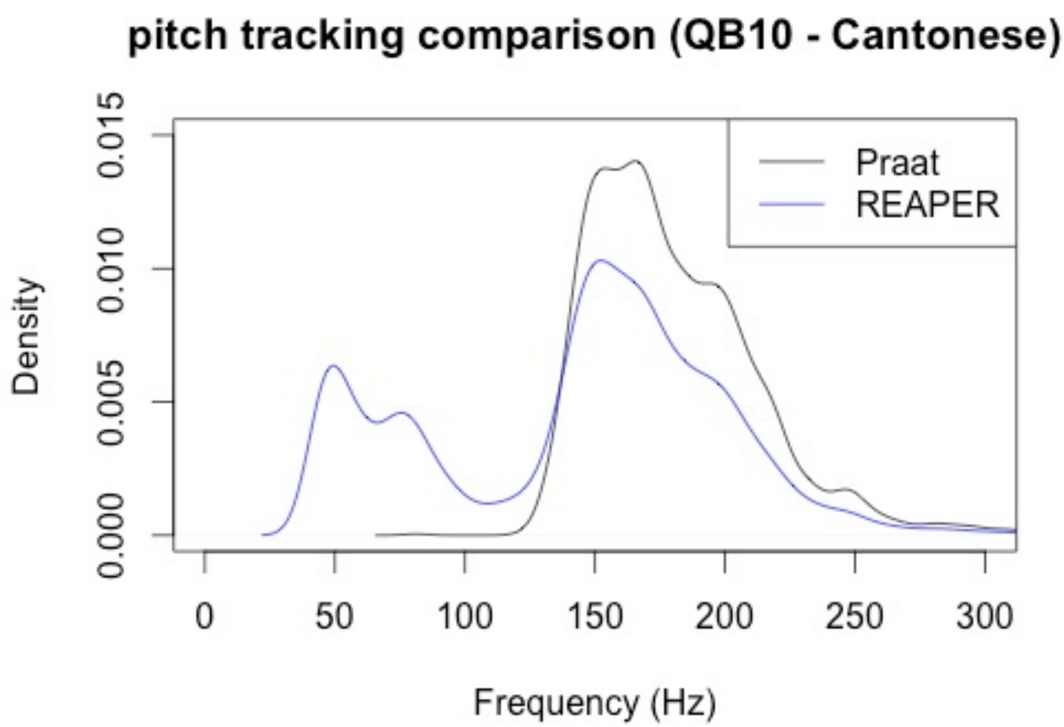


Figure 8 - Comparison of Praat and REAPER pitch trackers (Speaker QB10 - Cantonese)

Praat's pitch tracker was set to 40-400Hz for comparison with REAPER. It does not pick up on the lower frequencies as well as REAPER does (often misinterprets them as octave and fifth jumps), and will not track creaky glottal cycles as accurately as REAPER does. It is the greater number of creaky f0 measurements that pull down the mean f0 in REAPER. As a result, using REAPER provides a far more accurate total mean f0 due to its accuracy in measuring in the lower frequency creak range, with the caveat of this being a seemingly unrealistic, but substantially more accurate, result than seen in previous, Praat-based studies.

Figures 9 and 10 show the total mean values for MAONZE and QuakeBox speakers respectively, with each bar representing the languages of each speaker.

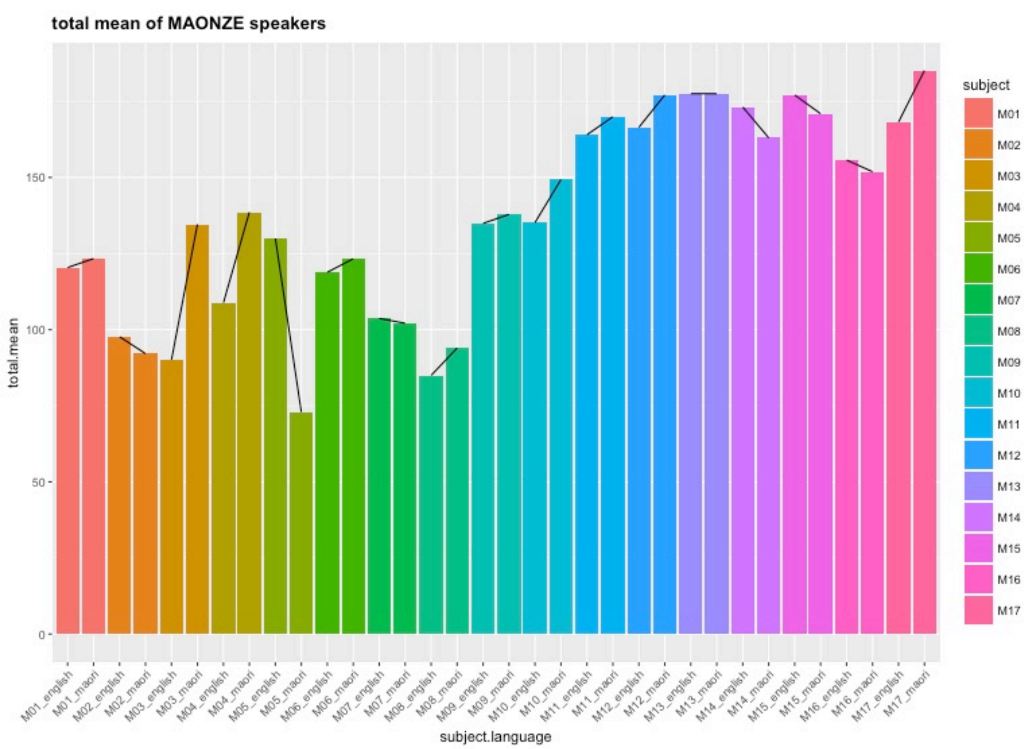


Figure 9 - Total mean f0 of MAONZE speakers

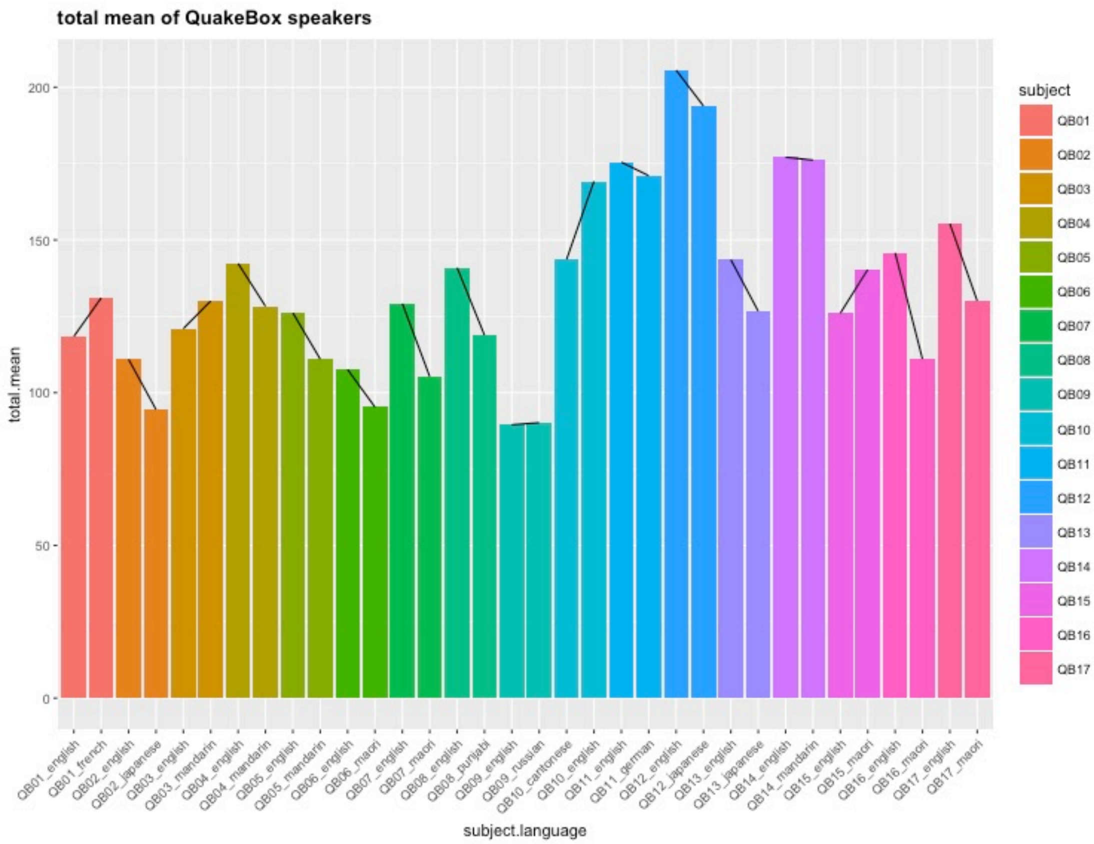


Figure 10 - Total mean f0 of QuakeBox speakers

The majority of the MAONZE speakers appear to be relatively similar when it comes to total mean f0. Speaker M13's total mean f0 for both languages is essentially equal, with speakers M01, M07, and M16 not far off equal most of the others (with the exception of speakers M03, M04, and M05) have total mean f0 values close together. This may give the illusion that total mean f0 is a reasonably good speaker discriminant. This would appear to be true if it wasn't for the effect of the amount of creak produced by each speaker. As demonstrated above, the amount of creak has the ability to shift the total mean f0 downwards. In the case of the MAONZE speakers the percentage of creak was near equal for both languages. Therefore, since the amount of creak was the same, there will be a near equal shift downwards for a speakers two total mean f0 measurements.

The QuakeBox corpus also shows the effect creak can have on the total mean f0. All the QuakeBox speakers showed differences in their creak percentages, with the majority of these differences being very large. Figure 10 shows that nearly all of the speakers show differences between their two total mean f0 measurements. These differences appear to correlate to the amount of creak a speaker uses in each language: the higher the percentage of creak, the lower the total mean f0 value. Unlike the MAONZE speakers, the QuakeBox speakers demonstrate that total mean f0 is not a good speaker discriminant as there is far too much variation within a speaker's two languages for a precise comparison.

The total mean f0 measurement is essentially the go-to f0 measurement for many forensic phoneticians, however its fluidity in its relationship to creak allows for inaccuracies. This coupled with the imprecise pitch tracking methods that have been prevalent in f0 studies, means that the total mean f0 values are essentially incorrect, or implicitly measuring only the modal phonation part of the probability distribution. Accurate pitch tracking tools such as REAPER ironically only heighten this inaccuracy, as they provide many more creak values than other, less accurate, pitch tracking tools. In order to fully capture the bimodal nature of pitch the two distributions must be considered separately instead of joined together. This will ensure that creak measurements cannot interfere with modal phonation measurements, and vice versa, creating a more accurate reading of the mean f0 and other parameters for measuring f0.

4.3.2 Creak Mean

In order to tease apart the two pitch distributions in speech creak phonation and modal phonation were separated at the first antimode of the distribution. This antimode is defined as the least frequently used glottal cycle between the two phonation types. Creak is considered to be all frequency measurements below the antimode. Figures 11 and 12 show the creak mean f0 values for speakers from the MAONZE corpus and the QuakeBox corpus, respectively.

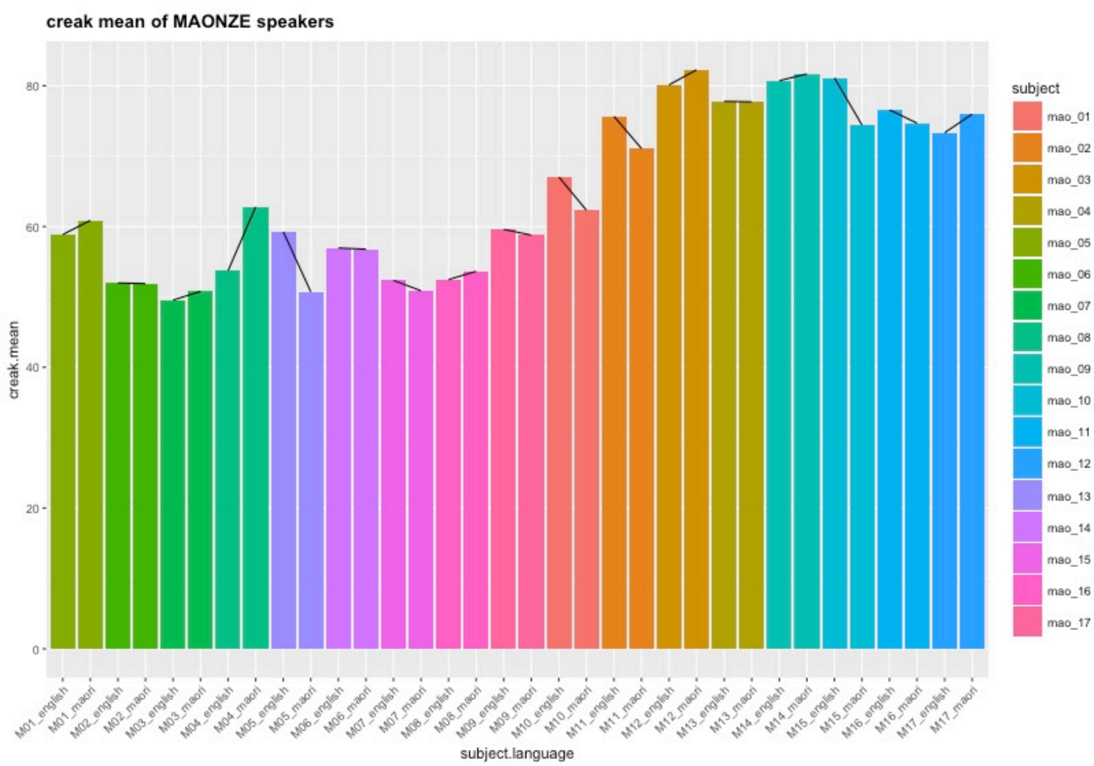


Figure 11 - Creak mean f0 of MAONZE speakers

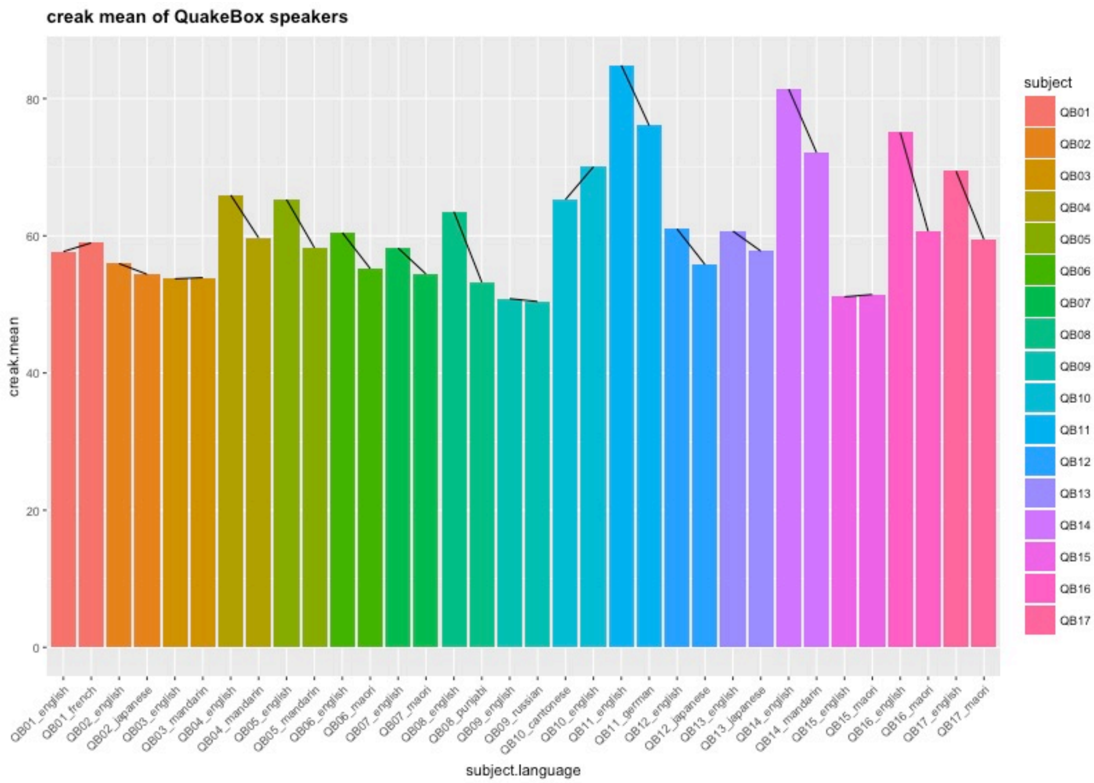


Figure 12 - Creak mean f0 for QuakeBox speakers

These figures continue the trend of the MAONZE speakers showing more stability within their two languages than their QuakeBox counterparts. All MAONZE speakers (bar speakers M04 and M05) have near equal creak means between their two languages. However, the range of creak means in the MAONZE speakers is relatively invariant meaning that discriminating speakers across the entire group based on this one feature is less successful. Male speakers show a lower creak mean f0 than female speakers, and the male speakers (excluding M04 and M05) show less variability with a smaller range of creak mean f0 (around 10Hz) than the female speakers, who show a range closer to 20Hz. However, the majority of female speakers fall into a relatively stable 10Hz range. Overall the majority of the MAONZE speakers show little to no variation within their creak mean values, however creak mean appears to be a relatively stable measurement across speakers, with each gender generally falling into a 10Hz range.

The QuakeBox speakers show much less creak mean variation within-speaker than they did with the total mean, but still show more variation than their MAONZE counterparts. However, there does appear to be a correlation between a speaker's two languages regardless of the amount of differences between them. The creak mean values of the speakers with the highest creak mean (speakers QB11 and QB14) appear to be relative to each other: QB11 has the highest English creak mean and the highest second language creak mean, and QB14 shows the same relationship as second highest. This appears to be consistent across all QuakeBox speakers; the difference between the two languages within each speaker is not minimal, however it is relatively independent across all speakers. This suggests the possibility that creak mean is speaker specific, albeit with a large range within-speakers. The pitch ranges for the QuakeBox speakers are far larger than those of the MAONZE speakers; male speakers have a range closer to 20Hz and female speakers are much more extreme with a range reaching past 30Hz. The larger pitch range for females is possibly due to the fact that that females generally have a higher pitch than males, so there is potentially more room for them to move around the pitch range. The antimode for females is located consistently higher than that of males, which gives females a larger acoustic space to produce creaky pitch.

Overall creak mean appears to be a more robust measurement for speaker discrimination than the total mean. There is much less variation as a whole within a speakers' two languages in creak mean than total mean, and there does appear to be evidence that the creak mean is speaker specific. However, the narrow range of attested mean creak values inhibits the ability to discriminate across speakers. Males especially show very similar mean creak values making it difficult to differentiate between-speaker with just the creak mean parameter.

4.3.3 Modal mean

The modal mean is the mean of the frequency of all modal phonation glottal cycles (i.e. values above the antimode). Figures 13 and 14 show modal mean values for each speaker's respective languages.

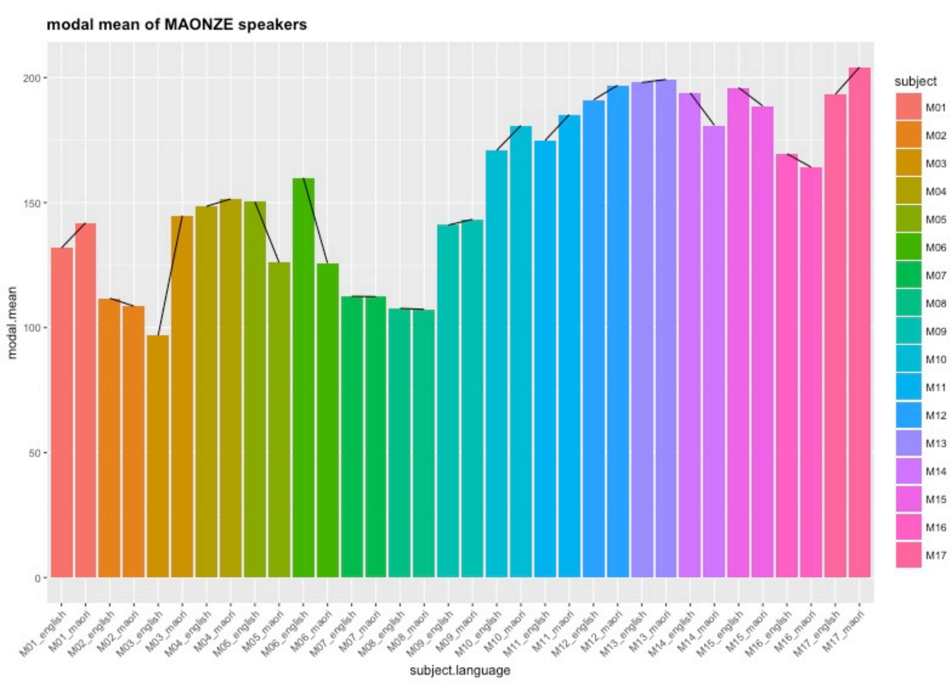


Figure 13 - Modal Mean f0 of MAONZE speakers

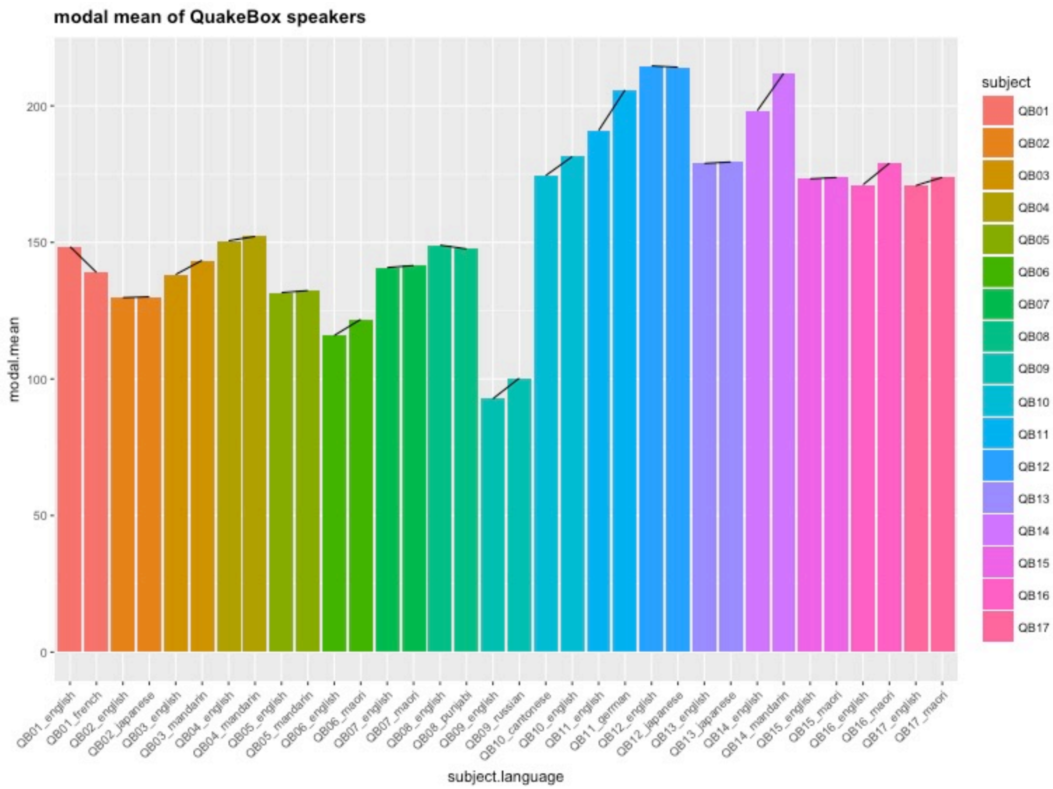


Figure 14 - Modal mean f0 of QuakeBox speakers

The modal means of the MAONZE speakers show similarities with their respective creak means. This means that for many of the speakers there appears to be a direct correlation between the creak mean values and the modal mean values of each speaker. Take for example speakers M15, M16, and M17. In figure 13 Speakers M15 and M16 both show a lower modal f0 in Maori than in English, while M17's values are higher. The differences seen between each speaker's f0 values is the same differences seen in their creak mean, as seen in figure 11. Speakers with little difference in their modal means (one in each language) also show little difference in their corresponding creak means. This is apparent in all speakers where their creak percentages (from section

4.2) are equal or near equal to each other, as in each speaker shows the same proportion of creak in both their languages. There are some outliers in this group, however: speakers M03, M05, and M06 all show large differences between their modal means. This is presumably because these three speakers do not follow the apparent norm of bimodality. While most speakers have one creak distribution and one modal distribution in each of their total f0 distributions, these three speakers do not. This will be discussed in more detain in chapter 5**.**

There is a stark contrast between QuakeBox speakers' modal means and creak means. While the QuakeBox speakers creak means showed little similarity within-speakers the modal means show much more similarity. Unlike all the other data presented thus far, the QuakeBox speakers modal means appear to be a better speaker discriminate than the MAONZE speakers. This is due to the QuakeBox speakers showing larger between-speaker spread and smaller within-speaker variability. The difference between the highest and lowest modal mean values for males is over 50Hz, and around 40Hz for females across both corpuses. This large range allows for more variability across the speakers which in turn allows for a more unique modal mean measurement in a way that creak mean could not serve. That is not to say the modal mean can achieve high speaker discrimination power all by itself; there are still speakers whose two modal means are not near equal. That, however, should not detract from the effectiveness of the modal mean as a speaker discriminant; it is so far the most idiosyncratic between-speaker parameter. There does not appear to be as strong a correlation between the QuakeBox speakers' creak means and modal means as seen in the MAONZE corpus. This is likely due to the more extreme differences in each speaker's percentage of creak.

Overall the modal mean by itself has shown to be a more capable speaker discriminant than that of creak mean and total mean. Apart from some outliers in the MAONZE corpus, each speaker's two (cross-language) modal means are near equal. This is especially evident in the QuakeBox corpus, which has up until this point shown much less similarities in the within-speaker measures. The larger between-speaker frequency ranges allow for more discriminatory power across speakers.

4.3.4 Total Mean vs. Split Mean

It is clear that the amount of creak a speaker produces can have considerable effect when the mean f0 is taken from both creak and modal distributions together. As tables 12 and 13 in appendix C shows, the average difference between a speakers two means is 12.98Hz for the total mean, 5.92Hz for the modal mean, and 4.19Hz for the creak mean. These numbers were calculated by finding the difference between the two means of each speaker then averaging the differences of all speakers. The outliers whose speech displayed a distribution type different from the vast majority's bimodal distribution (i.e. unimodal and trimodal distributions) were excluded from this calculation. These numbers show just how ineffective the total mean measurement is in displaying underlying speaker identity, with the average total mean difference being over double the average modal mean difference. The average creak mean value is also much lower than the total mean average, however it has been shown that there is a much tighter range in the creak mean parameter, so a lower average difference is to be expected.

Overall, the splitting of frequency distributions into the two phonatory ranges shows a more accurate representation of a speaker's mean f0. However, the means are not impervious to the effects of the percentage of creak. The QuakeBox speakers show that if there is a larger within-speaker difference in creak percentage there is a larger within-speaker difference in the creak mean. This appears to only affect the creak mean, as the modal means of the QuakeBox speakers do not seem to show this. The same effect can be seen but to a lesser degree in the MAONZE speakers.

Splitting the total mean into two phonatory distributions also allows for patterns between the two distributions to emerge that cannot be seen when not separated. If, for example, a MAONZE speaker shows a higher creak mean in their English recording than their Maori recording, then the same appears to be true for the modal mean. This however is also dependent on a speakers creak percentage, as this correlation only seems to show when a speakers creak percentage is near equal between languages.

Overall splitting mean f0 into its two frequency distributions paints a far more nuanced and accurate picture of the mean f0 than simply calculating the total mean f0 over (often only a portion of) the whole phonatory range does. However, the effectiveness of mean as a speaker discriminant does appear to be limited to the percentage of creak a speaker has, as the percentage of creak has a significant effect not only on the total mean, but also on the creak mean. In this study the whole measurable frequency range down to 40Hz has been used to track the speakers pitch, something which most previous studies do not do. This is usually due to using a pitch tracker that cannot manage creak. Usually an arbitrary pitch range will be implemented on to both ends of a speaker's frequency distribution, which removes entire sections of a speaker's f0 measurements. A mean value from 50-200Hz is going to be different to a mean value of 40-400Hz. Gold (2014) is a good example of this, as she manipulated the pitch floor and pitch ceiling values for multiple speakers in her study. Because REAPER was able to accurately track a speaker's entire frequency range down to 40Hz there is a correlation between the pitch tracker range and what a speaker actually produced, which allows for a more overall transparent set of measurements.

4.4 Mode

The mean f0 is a reasonable speaker discriminant but is far too malleable when influenced by other parameters, namely the percentage of creak a speaker uses. A more ideal parameter would be one that matches the accuracy seen with the mean f0 in both within-speaker and between-speaker but is not affected by other parameters like creak percentage and outliers. The mode f0 is a relatively robust measurement that potentially fits within these guidelines. As mode f0 is simply the most frequently used pitch in a distribution it often does not change when other parameters change. Where the total mean f0 moved along a speakers distribution dependent on the amount of creak they produce, the mode f0 stays static. This fact alone shows that the mode f0 may be a more robust measurement than the mean f0. But how does it hold up as a speaker discriminant? Does it show the same within-speaker and between-speaker patterns that the mean f0 showed? To ascertain this the creak mode and modal mode for each MAONZE and QUAKEBOX speaker's recordings were obtained, with the cut off point for creak and modal phonation being the first antimode from each recording.

4.4.1 Creak Mode

Figures 15 and 16 show the creak mode values for the MAONZE and QuakeBox speakers, respectively.

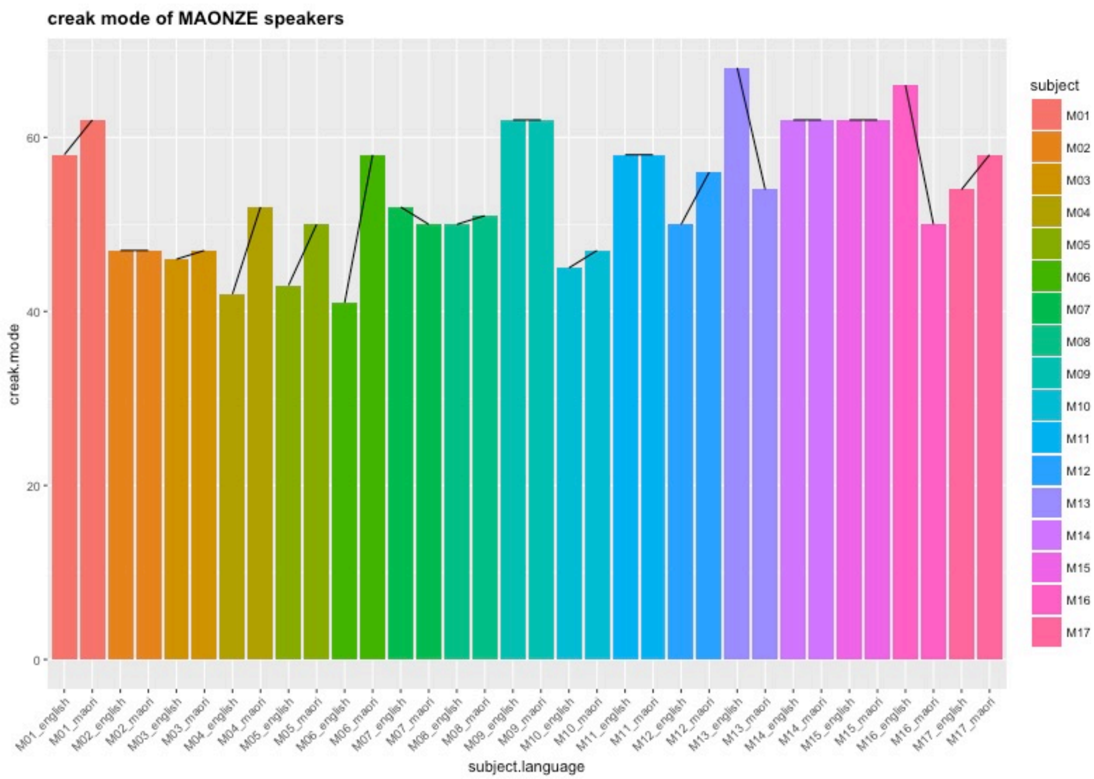


Figure 15 - Creak mode of MAONZE speakers

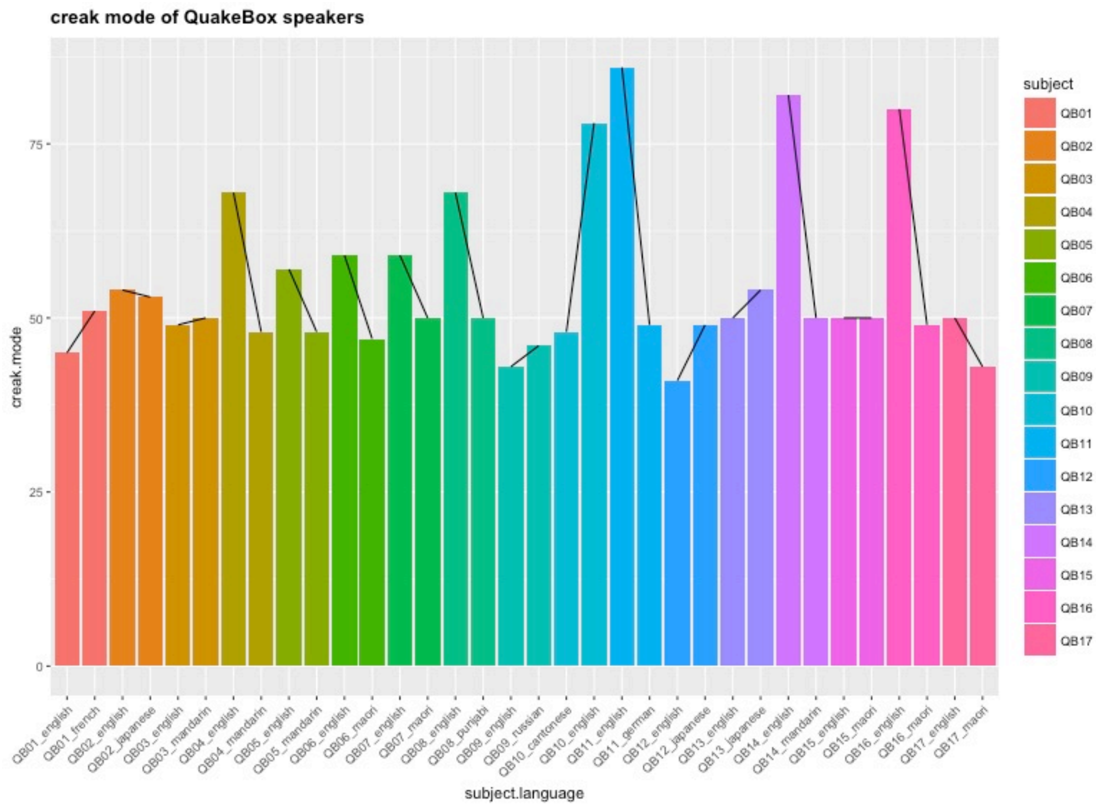


Figure 16 - Creak mode f0 of QuakeBox speakers

The creak mode does not appear to be as good a speaker discriminant as creak mean was. Many speakers in the MAONZE corpus show equal or near equal creak modes in their two languages, however there are still many speakers who show large within-speaker differences. Granted, some of these are speakers who have been consistently irregular through all parameters (such as the unimodal and trimodal speakers), but others (such as speakers M13 and M16) show large differences in their creak mode while their creak mean is near equal. However, as a whole similar creak modes are seen within the MAONZE speakers two languages. The same cannot be said about the QuakeBox corpus.

While there are a few QuakeBox speakers who show similar or near equal creak modes (such as speakers QB02, QB03, and QB15), the majority shows a significant degree of separation within-speaker (across languages). The most notable examples are speakers QB11 and QB14, who show a difference of around 30Hz each. The creak mean did show similar results in that there were marked differences between each speakers languages, however the creak mode differences are far more extreme than those seen in the creak mean.

With the creak mean there was a small but clear separation between males and females, with the females consistently having a higher creak mean than males. This separation is not so clear within the creak mode. The MAONZE female speakers appear to have a slightly higher creak mode, however there are two male speakers (M01 and M09) that are equal to, if not higher, than the female speakers. The creak mode of the QuakeBox speakers is very stable. Looking at only the lowest creak mode for each speaker there does not appear to be a clear differentiation between male and female. Females that have large differences in their creak modes tend to have much larger differences than males do, and this appears to b e the only way to distinguish between the two genders of this corpus using the creak mode. However, this is not consistent across all female speakers. Some of them, such as speaker QB15, are near equal and can be directly compared with a male speaker like QB03. Interestingly the speakers with the lowest creak mode (QB12 and QB17) are both female speakers. This just exemplifies the inconsistencies that the creak mode parameter holds.

4.4.2 Modal Mode

Next the mode of the modal phonation is explored. Figures 17 and 18 show the modal mode values for the MAONZE speakers and QuakeBox speakers, respectively.

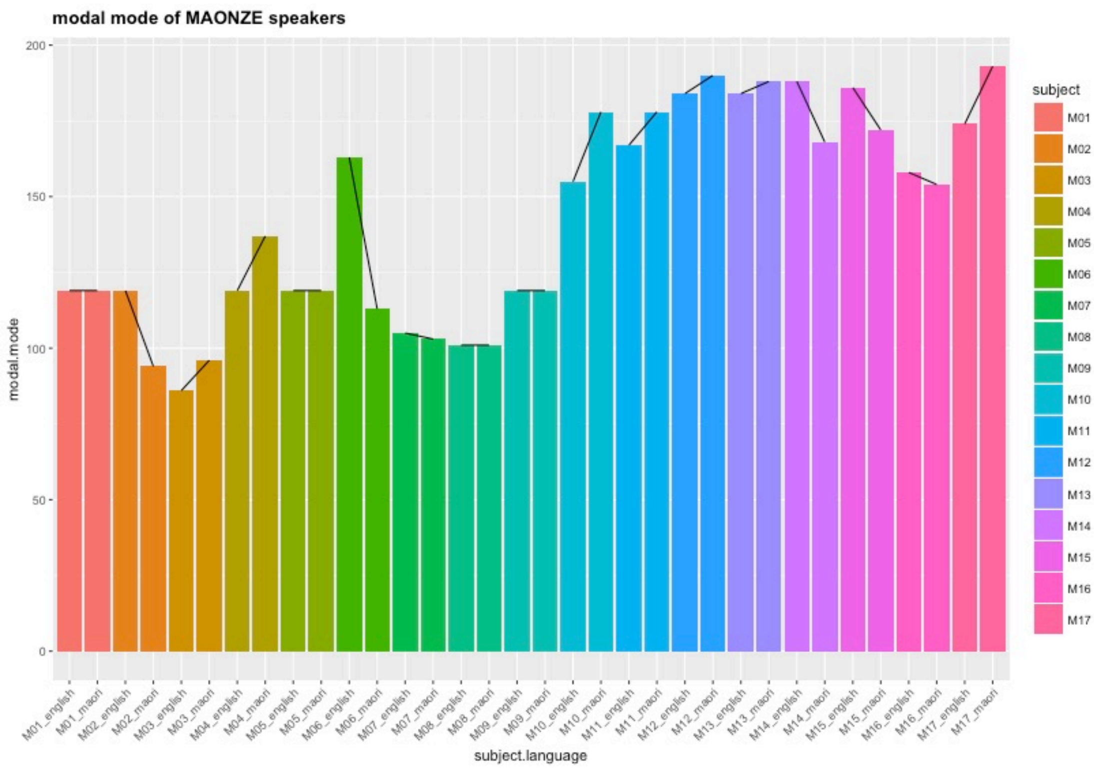


Figure 17 - Modal mode f0 for MAONZE speakers

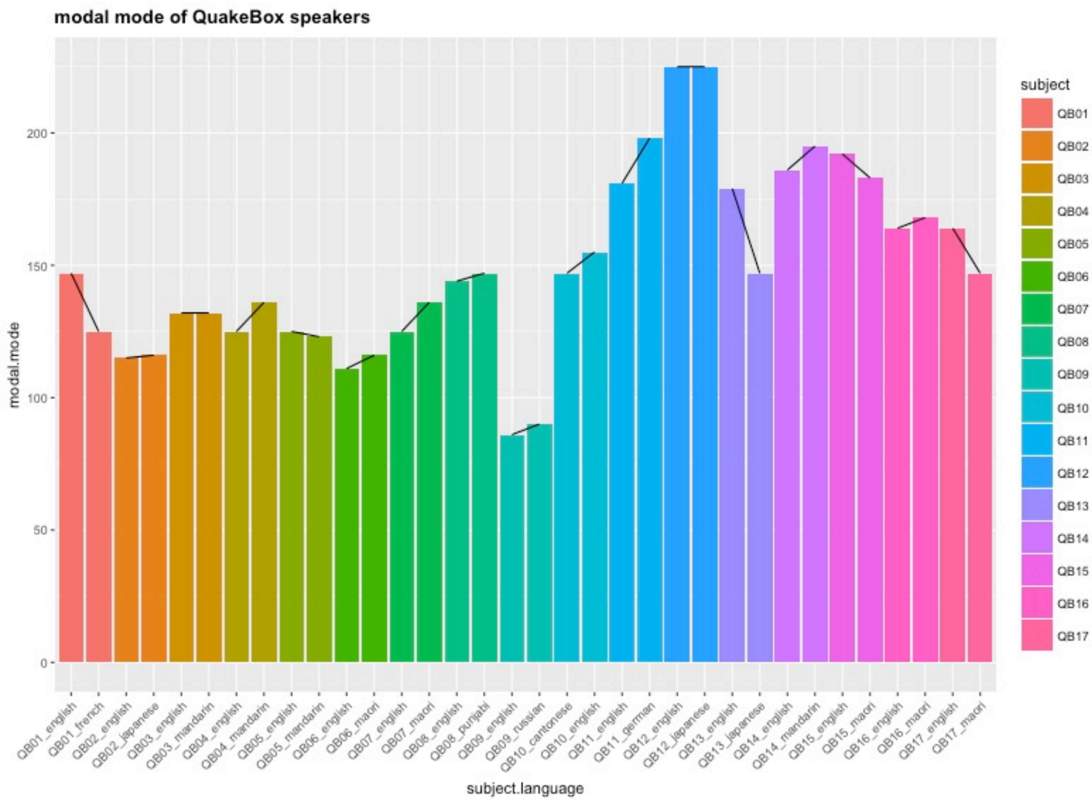


Figure 18 - Modal mode f0 for QuakeBox speakers

There is a stark contrast between the modal mode and the creak mode, as the modal mode shows reasonably small within-speaker variation and a large between-speaker spread. The majority of the speakers in both corpuses show either equal or near equal modal mode values between their two languages, similar to the results seen in the modal mean values. The patterns within-speakers are also similar in the modal mode as they were in the modal mean; if a speaker has a higher modal mean in English than their second language, the same generally applies for their modal mode. The average difference between a speaker's two languages for modal mode is 10.19Hz (appendix C, tables 12 and 13), which is nearly double that of the modal mean. So while the modal mean and the modal mode act similarly, the modal mean does appear to show a smaller range within a speaker.

The modal mode, however, does have a reasonably large frequency spread between-speakers. The male MAONZE speakers (excluding the M06 outlier) have a frequency range of around 50Hz, while the MAONZE females show a slightly smaller range at around 40Hz. The males do appear to be more stable within that range, with the females showing more pitch variation both within themselves and between each other. The male QuakeBox speakers are similar to the MAONZE males, with a range of around 50Hz but with modal mode values reasonably stable within that range. The female QuakeBox speakers show a much larger frequency range than both males and the MAONZE females at around 75Hz. The QuakeBox females are also not as stable as their male counterparts, with the speakers showing much more variability within-speaker and more spread between-speakers.

Overall the male speakers are much harder to differentiate from each other, as their modal mode values are relatively similar. However, when comparing each speaker's two languages the male speakers show greater similarities than the females. It is clear that the modal mode is better at discriminating speakers than creak mode, with modal mode showing a higher rate of similarities within-speakers and a reasonably high amount of between-speaker spread, especially among the females. There does not appear to be any correlation between the creak mode and the modal mode, an effect that was seen between the creak mean and modal mean (cf. Section 4.3).

4.4.3 Mean vs. Mode

It has been shown that both mean and mode may function as reasonable speaker discriminants. They both have relatively little variation within a speaker, yet are quite variable across speakers. But which one is the better measurement? Does mean f0 or mode f0 provide a more accurate picture of a speaker against the backdrop of a larger group of speakers?

Predictably, the modal mode and the modal mean do relate to each other within a speaker. If a speaker's first modal mean value is higher than the second the same generally appears to be true for the modal mode. The modal mean does show a smaller average distance than its mode counterpart, which has on average nearly double the amount of difference. This isn't to say the modal mode is inherently worse than the modal mean; on an individual speaker level both mean and mode show reasonably similar between-speaker spread. The modal mode just shows less stability within a speaker, especially in female speakers. When looking across all speakers, there does seem to be some reasonably large variation through both groups of speakers, with females showing greater variability than males. The between-speaker variation is reasonably similar in both the mean f0 and the mode f0, however it appears that the mode f0 performs marginally better than the mean f0 as it has a larger between speaker spread than the mean f0 in both creak and modal phonation.

There does appear to be a correlation between the creak mean and the modal mean, similar to the relationship between modal mean and modal mode. If a speaker's first creak mean value has a higher frequency than the second the same generally appears to be true for the modal mean. However, this relationship only materializes when the percentage of creak for a speakers two languages are equal or near equal. This correlation is unique to the mean f0 and it is not present in the mode measurements.

The creak mean/modal mean correlation highlights one of the main drawbacks of the mean f0 parameter, that being the high level of influence the amount of creak a speaker uses has on the overall values. The mean f0 has the potential to move around the total frequency distribution of a speaker, depending on factors such as how much creak a speaker produces, or their pitch variability within a frequency distribution. The mode f0 on the other hand is a lot more stable. Since the mode is the most frequently used pitch value, the only way for it to change is for a speaker to use another pitch value at a higher frequency. In other words it is impervious to influencing factors in a way that mean f0 is not.

Overall neither parameter is a perfect speaker discriminant in and by itself as they both have their drawbacks that affect their performance. The mean f0 does work better than the mode f0 when there are no influencing factors, as it shows greater similarities within speakers and also correlations between the two means in a bimodal distribution. This is useful for identifying a speaker through multiple parameters. However, given the variability of the amount of creak a speaker uses there is a lot of opportunity for error when calculating the mean f0. The mode f0 is a very stable measurement, but does not show any correlation between creak mode and modal mode, and shows fewer similarities within a speaker than its mean counterpart.

4.5 Skew and Kurtosis

Skew and kurtosis are measures of asymmetry in a distribution, with skew showing the level of differences either side of the mean and kurtosis showing the level of peakiness of a distribution. These measurements have been shown to have the potential to be reasonable speaker discriminants (Rose, 2003; Kinoshita et al., 2009). However, these measurements are only easily interpretable in unimodal distributions. Most studies that look at skew and kurtosis (such as Kinoshita et al. (2009)) have done so disregarding the bimodal nature of pitch. Like the previous sections on mean f0 and mode f0, this section will look at the skew and kurtosis of creak phonation and modal phonation separately. This will allow potential patterns between creak phonation and modal phonation to emerge, and will provide more accurate skew and kurtosis measurements than when creak phonation and modal phonation are treated as one.

4.5.1 Creak Skew

Figures 19 and 20 show the skewness of the creak mode for MAONZE and QuakeBox speakers respectively. Most of the speakers in both corpuses show a positive skew; the distributions were weighted more on the right of the creak mode. Four speakers in the MAONZE corpus showed negative skew (M06, M09, M11, and M16), though out of these four only one speaker (M16) was consistently skewed to the left in both of their recordings. Males tend to have a higher positive skew, and are much more variable in the amount of skewness. The females on the other hand show skewness close to zero, which indicates that their creak distributions are far more equal. One female (M12) did show a completely equal creak distribution, showing a skewness of zero in their Maori speech.

Speaker M05 was omitted from figure 19, as their positive skew of around 4 was a clear outlier.

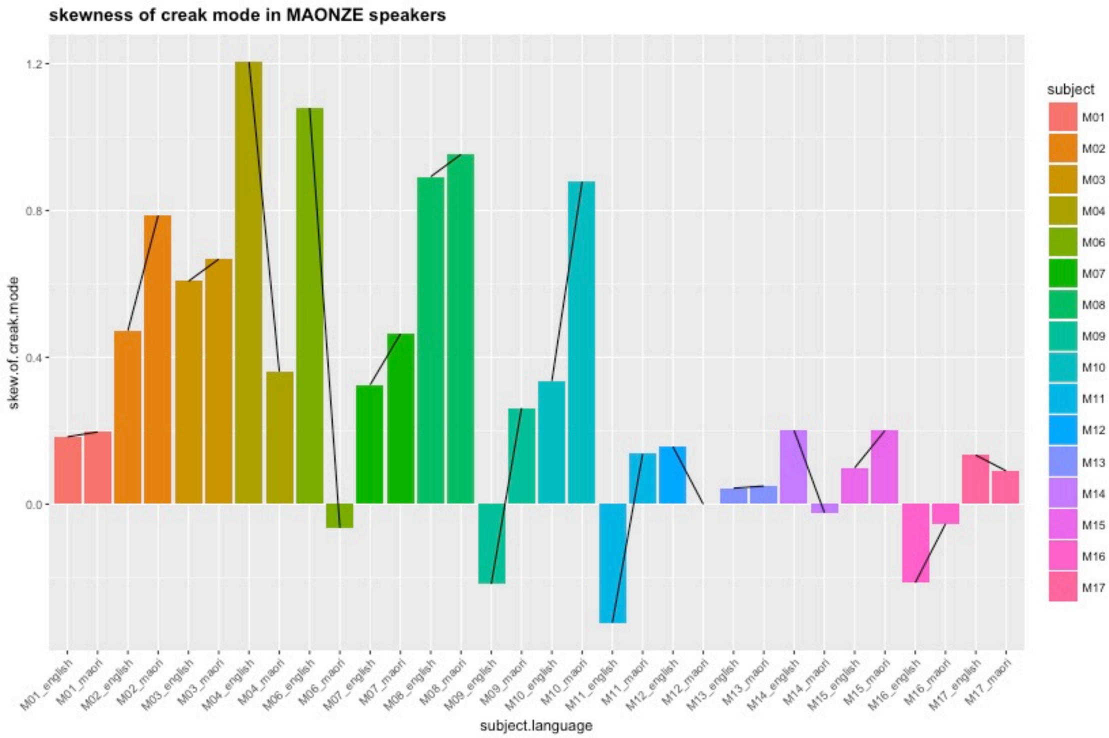


Figure 19 - Skewness of creak mode in MAONZE speakers

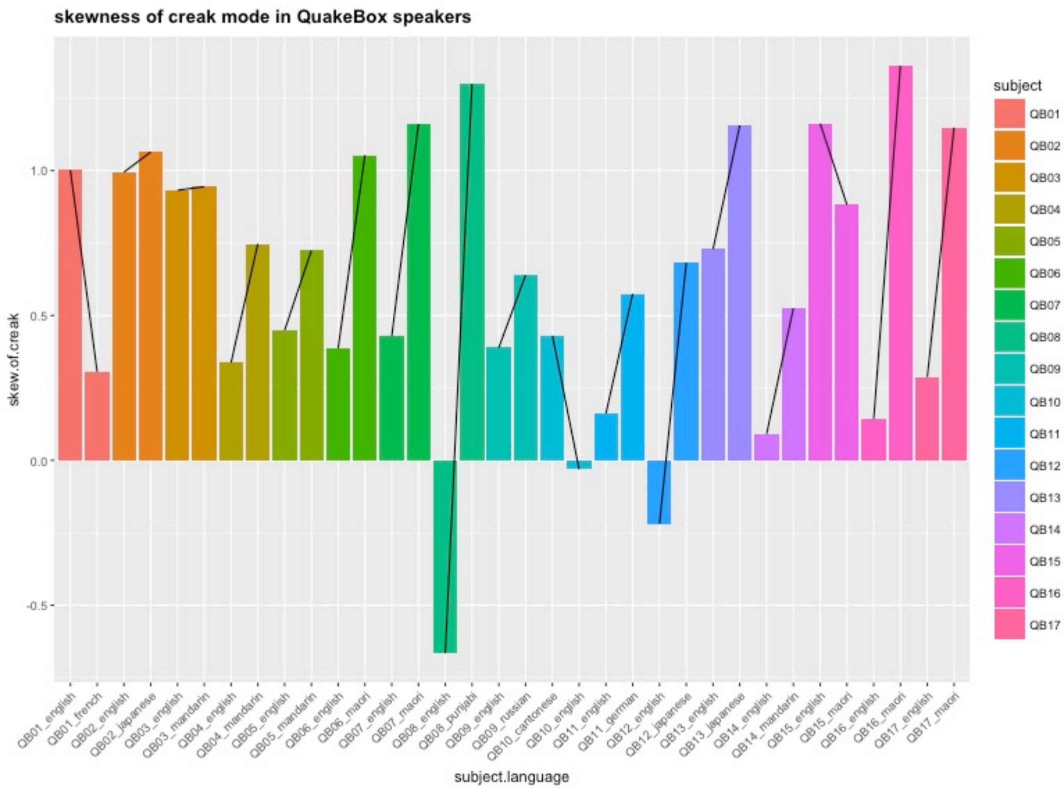


Figure 20 - Skewness of creak mode in QuakeBox speakers

The QuakeBox speakers show less systematic measures of skew, but most speakers show positive skew in both their languages. Only three speakers (QB08, QB10, and QB12) show negative skew in one of their languages, with QB10's English coming very close to equal distribution. No speaker from the QuakeBox corpus had consistent negative skew in both languages. Unlike the MAONZE speakers, there is no gender difference for the QuakeBox speakers. Both males and females act similarly with reasonably large differences between their creak skew values, very few of which come near to equal distribution.

4.5.2 Modal Skew

Figures 21 and 22 show the skewness of the modal mode for the MAONZE speakers and QuakeBox speakers respectively. All speakers in the MAONZE corpus show positive skew for each recording, but there isn't a strong link between the amount of skewness in a speaker's two languages. Some speakers, such as M02 and M12, appear to be reasonably close in their respective skewness, but this does not appear to be the norm. The amount of skew is also much larger than that of creak, this is to be expected as modal phonation is generally more common and has a bigger range than creak phonation. There is no gender difference in the MAONZE speaker's modal skewness, both genders appear to behave similarly.

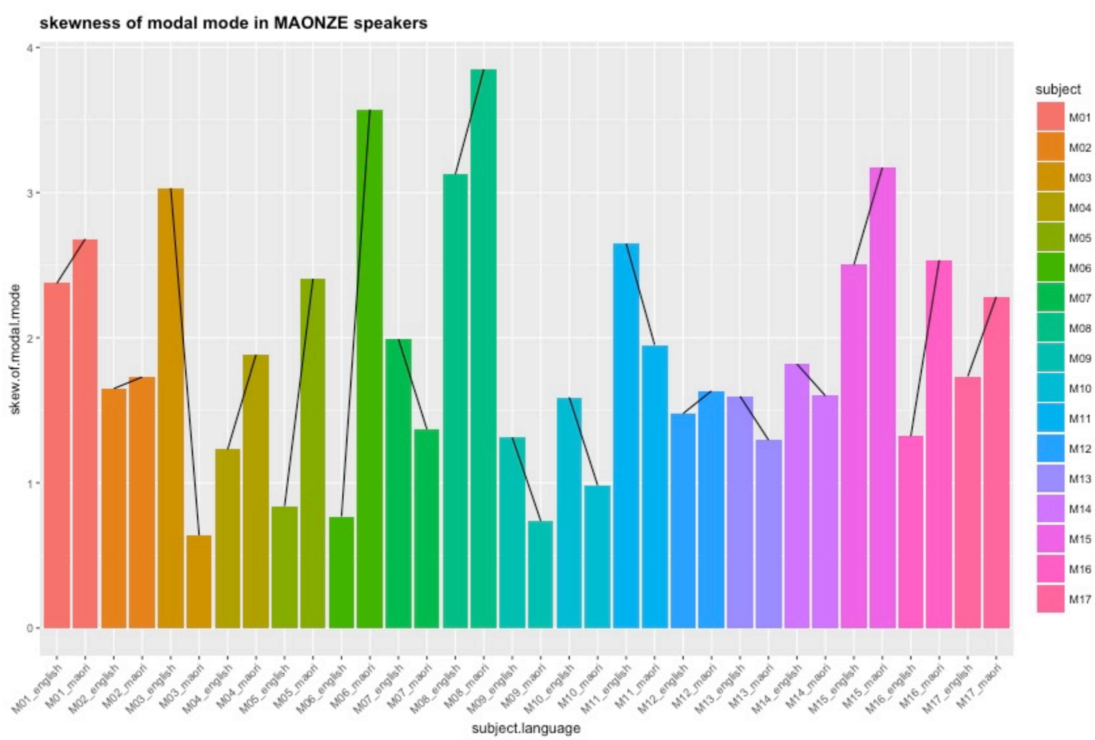


Figure 21 - Skewness of modal mode in MAONZE speakers

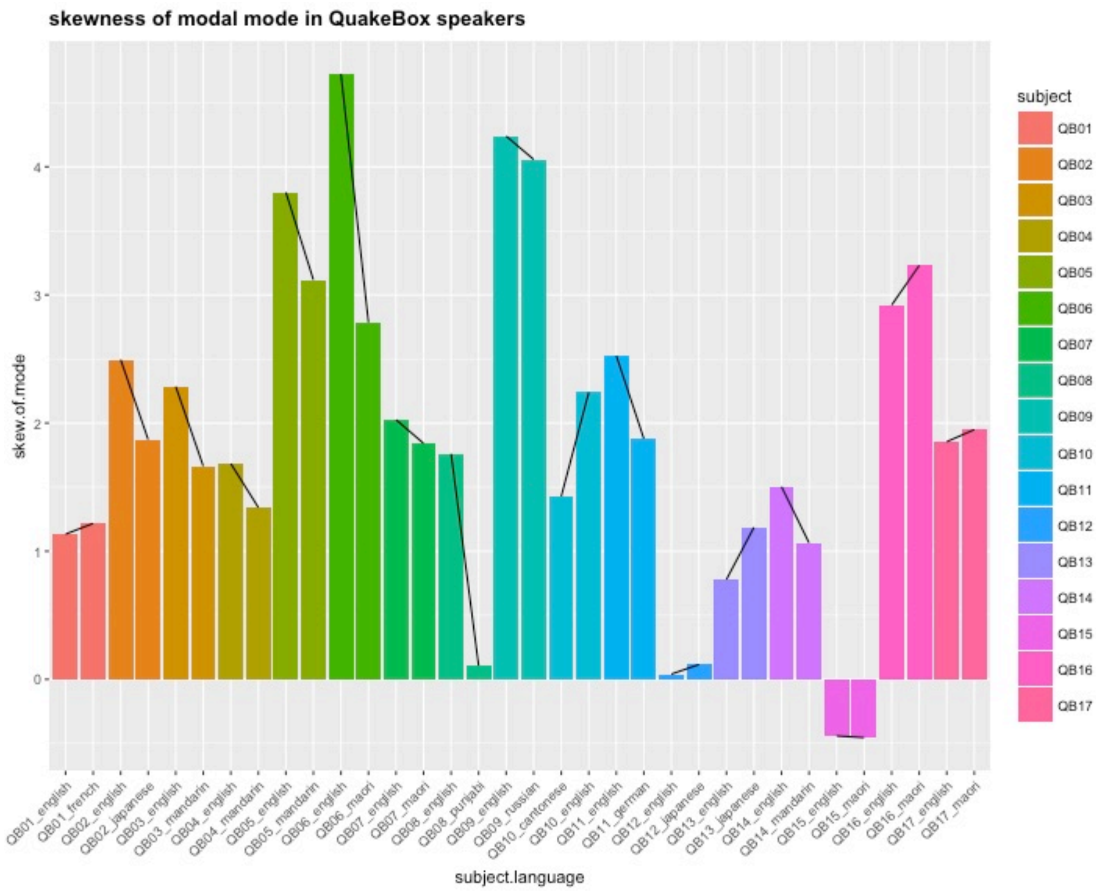


Figure 22 - Modal skewness for the QuakeBox speakers

The QuakeBox speakers appear to perform in much the same way as the MAONZE speakers, in that their skewness is generally positive with little similarities within a speakers two languages. There are three speakers from the QuakeBox corpus that approach zero (or equal distribution): QB08, QB12, and QB15. QB08 approaches equal distribution in only one language (Punjabi), while Japanese speaker QB12 approaches zero in both English and Japanese. QB15 is the only speaker in both corpuses to show negative skewness of the modal mode, and is also the speaker who shows the most similarity between their two skewness values.

4.5.3 Creak Kurtosis

Like skewness, kurtosis is measured with zero being the equal distribution. The further away from zero a kurtosis value gets the more imbalanced the distribution is. A positive kurtosis is called leptokurtic, which signifies a higher peak with skinnier tails. A negative kurtosis is platykurtic, which is a distribution with a low peak and heavy tails.

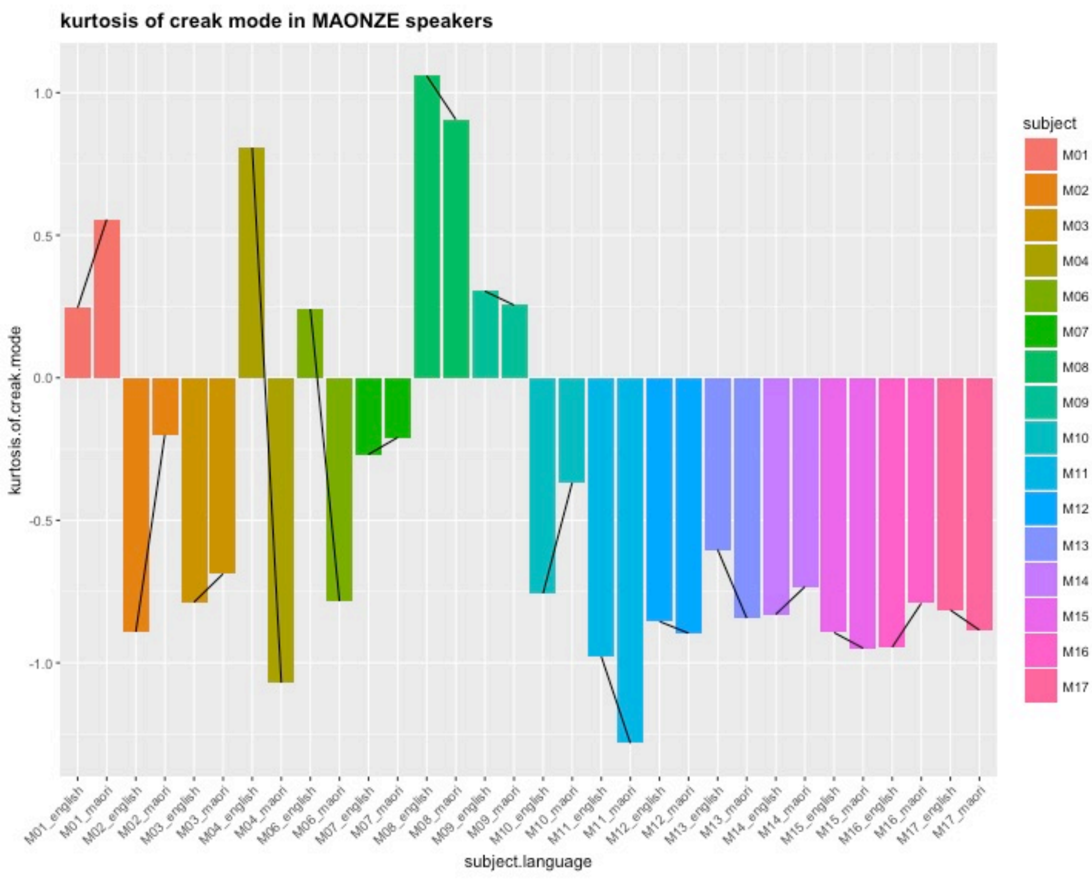


Figure 23 - Kurtosis of creak mode in MAONZE speakers

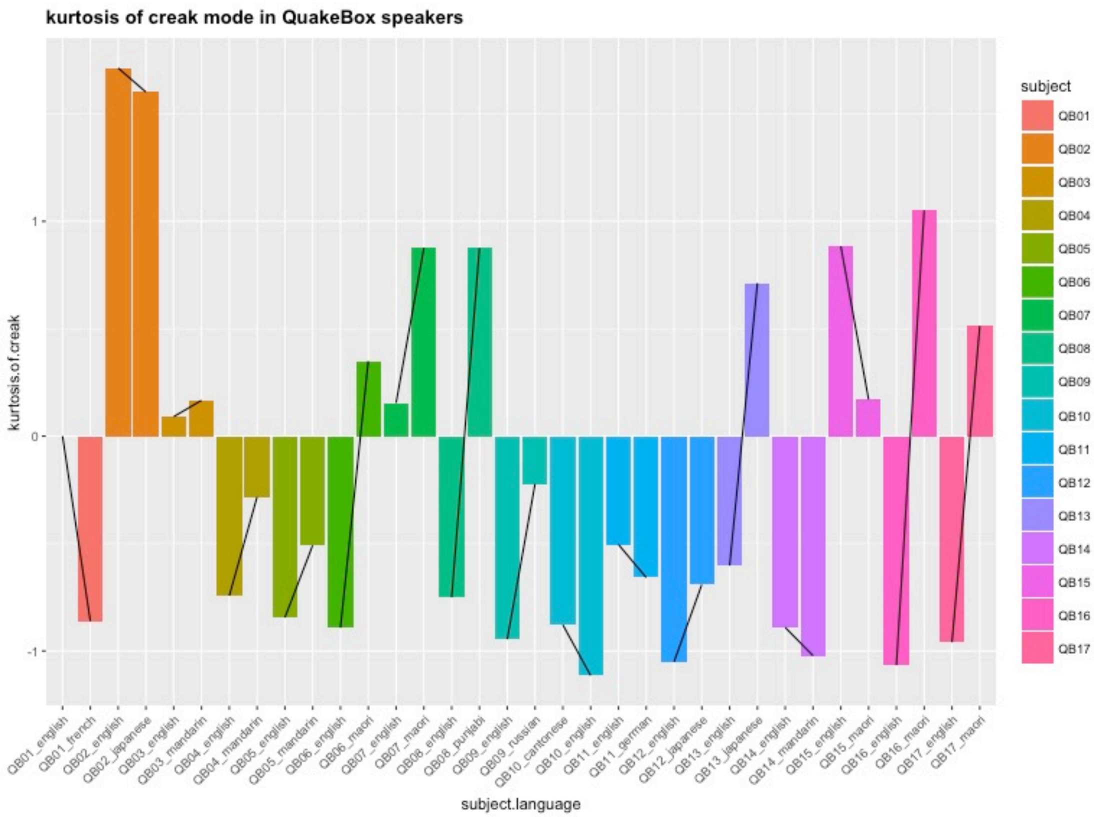


Figure 24 - Kurtosis of creak mode in QuakeBox speakers

Figures 23 and 24show the kurtosis measurements for the MAONZE speakers and the QuakeBox speakers respectively. M05 was omitted from figure 23 as his English kurtosis value was approaching 20, a clear outlier considering the rest of the MAONZE speakers are all relatively close to zero. The majority MAONZE speakers show platykurtic distributions, however there are a few exceptions. M04 and M06 both have one of each type of kurtosis, with their English being leptokurtic and their Maori being platykurtic. M01, M08, and M09 are all exclusively leptokurtic. Compared to the skewness of the creak mode, the kurtosis measurements for each speaker are relatively close together. Only the males have leptokurtic distributions in either one or both of their frequency distributions, the female MAONZE speakers are exclusively platykurtic.

The QuakeBox speakers show much more variability in their creak mode kurtosis. Over half the speakers show leptokurtic distributions in at least one of their frequency distributions, which is split relatively evenly across genders. Only four speakers (QB02, QB03, QB04 and QB15) are exclusively leptokurtic. There is one speaker who shows equal distribution in their creak mode: QB01's English kurtosis value is at zero.

While it is clear that no speaker in either corpus shows a completely equal distribution, most speakers' kurtosis measurements are relatively close to zero. This means that while they are not equal distributions, they are very close to it.

4.5.4 Modal Kurtosis

Figures 25 and 26 show the modal mode kurtosis values for the MAONZE speakers and QuakeBox Speakers respectively. The modal mode kurtosis for both corpuses is predominantly leptokurtic. This is contrastive with the creak mode kurtosis, where platykurtic distribution appeared to be the majority kurtosis distribution. Another stark contrast between creak mode kurtosis and modal mode kurtosis is the range of values given. Most creak mode kurtosis values lie between one and negative one. The modal mode kurtosis range for the MAONZE speakers is between just below zero and 20. The QuakeBox speakers are even more extreme with one speaker (QB09) approaching 50. The large values indicate that there is a very tall peak with skinny tails, meaning that the speaker has a very limited range of pitch. This appears to be more common in male speakers, who generally have the largest modal mode kurtosis values. It is however not unique to males, female speaker QB16's kurtosis values approach 20.

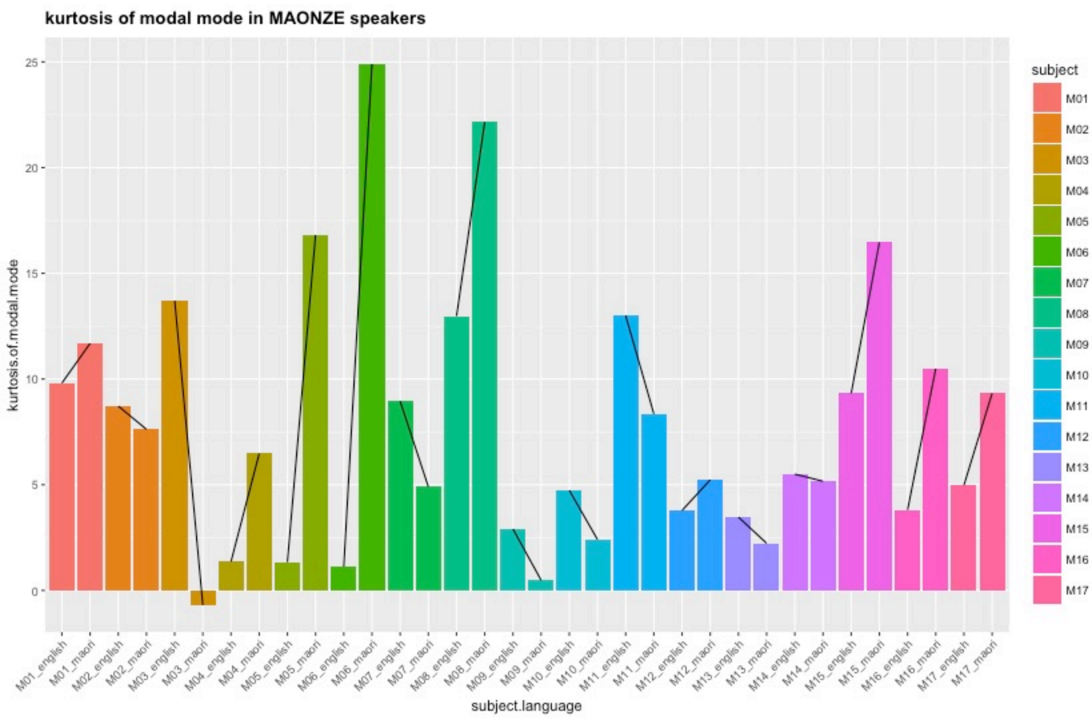


Figure 25 - Kurtosis of modal mode in MAONZE speakers

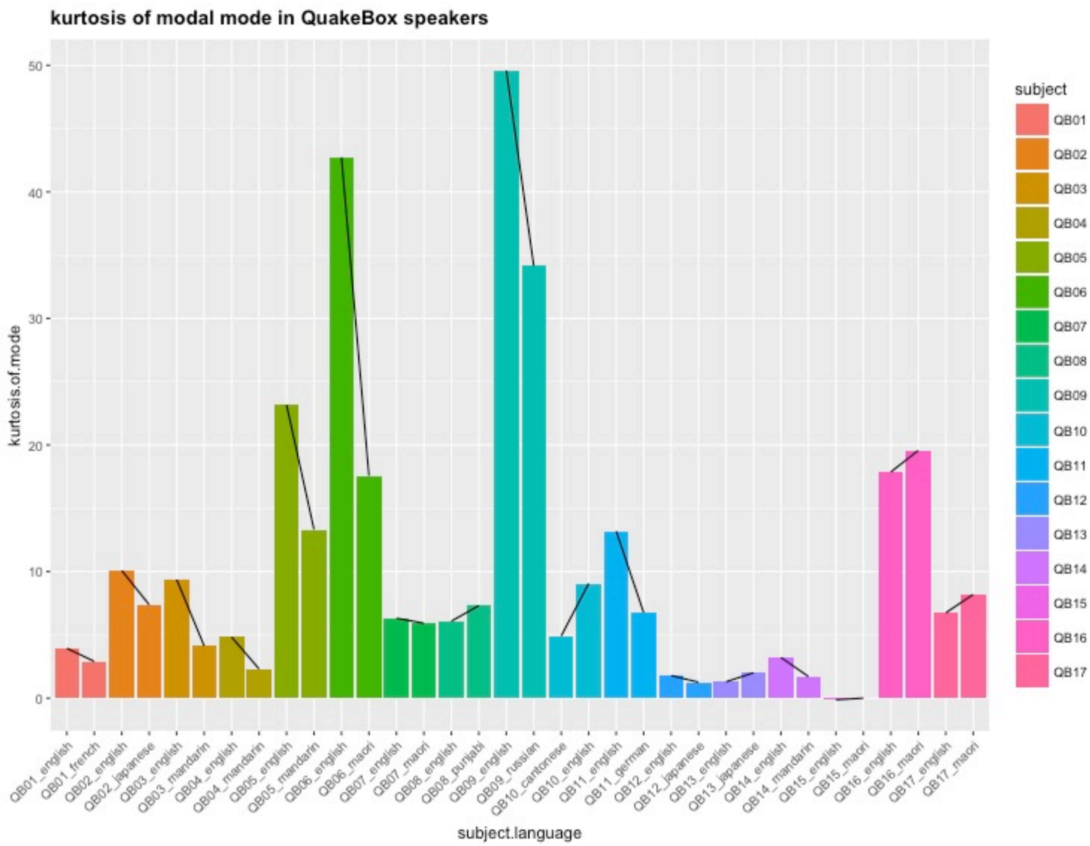


Figure 26 - Kurtosis of modal mode in QuakeBox speakers

Platykurtic distribution appears to be very rare in the modal mode kurtosis, but not impossible. Speaker M03’s Maori kurtosis value is below zero, as is speaker QB15’s English kurtosis value. However, both these values are very close to zero, especially QB15’s.

4.5.5 Conclusions on Skew and Kurtosis

Skew and kurtosis do not show much in the way of within-speaker variability or between-speaker spread. This applies for both creak phonation and modal phonation. No patterns between skew and kurtosis are evident, there does not appear to be any way to connect the two as measures of the same distribution. There also does not appear to be a way to connect a speaker's two language disributions using either skew or kurtosis. This shows that the shape a speaker's individual language distribution makes is wholly independent and does not correlate between languages. There also does not appear to be a connection between the skew and kurtosis of a speakers creak distribution and modal distribution.

Overall there is no evidence in the data to suggest that skew and kurtosis are effective speaker discriminants. They are too variable within a speaker (creak distribution and modal distribution), across a speaker's languages, and over a group of speakers.

4.6 Antimode

The antimode is the opposite of the statistical mode; it shows the location of observations that occur with the lowest frequency between two modes. This parameter was originally calculated in this study as the point of separation between a speaker's creak phonation and modal phonation. During this process it quickly became apparent that not only was there a large range of antimode values, it was also very similar within each speaker. This small within-speaker variability and large between-speaker spread is ideal for speaker discrimination purposes (cf. chapter 1, p. 1). Figures 27 and 28 show the antimode values for each speaker of the MAONZE corpus and QuakeBox Corpus respectively.

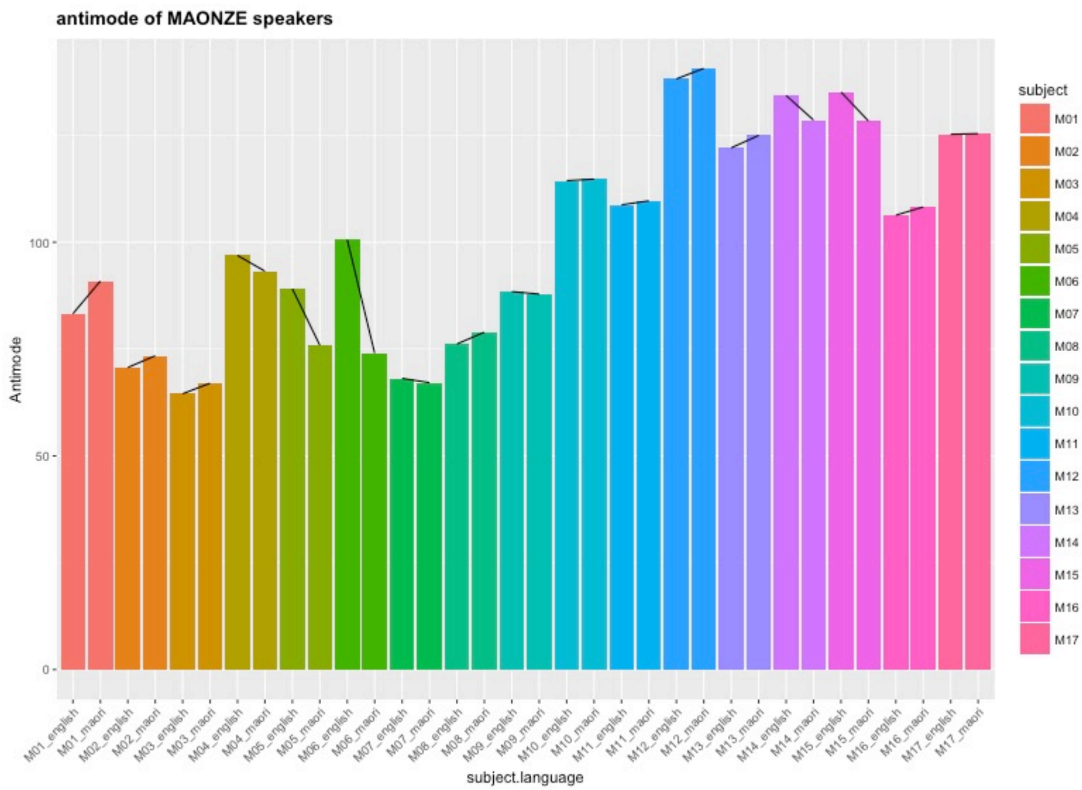


Figure 27 - Antimode of MAONZE speakers

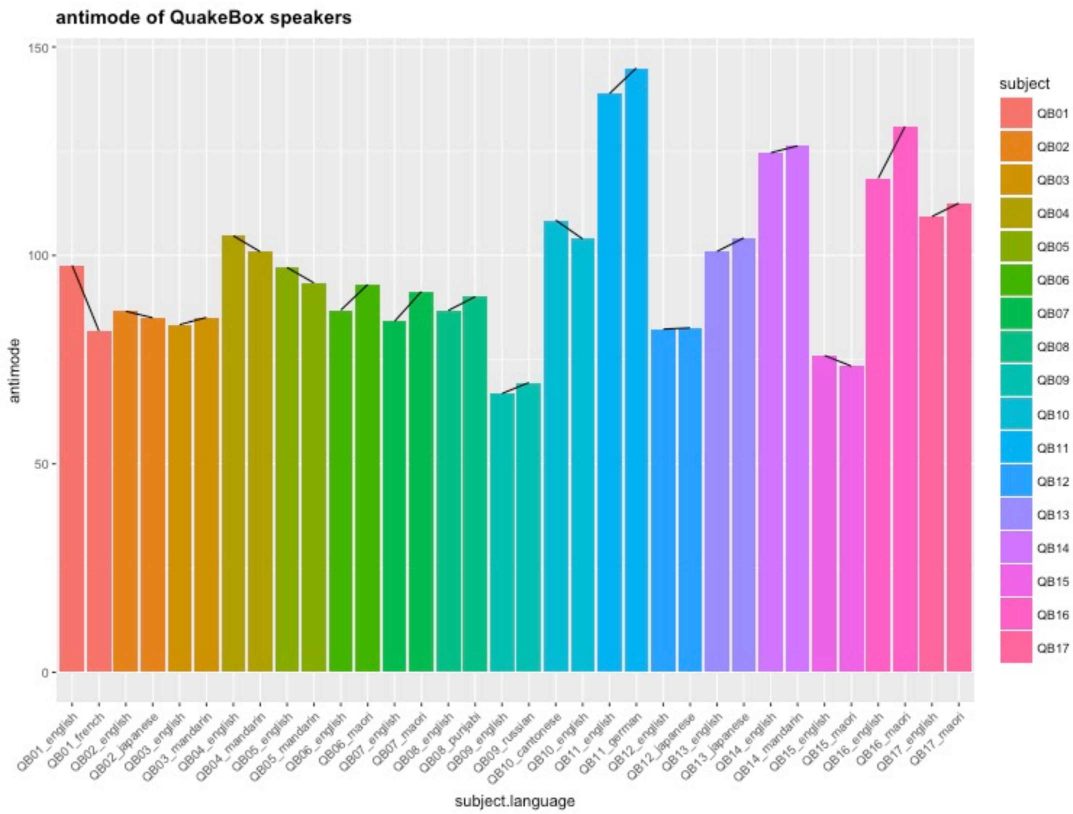


Figure 28 - Antimode of QuakeBox speakers

Nearly all the speakers of the MAONZE corpus show near-equal antimode values. The exceptions to this are the M05 and M06 outliers (these outliers will be discussed further in Chapter 5). Five speakers (M07, M09, M10, M11, and M17) show little to no within-speaker variation, with the rest coming very close. M01, M14, and M15 are the non-outlier speakers who show the most within-speaker variation, however they are still reasonably close to each other. There also seems to be a reasonably large between-speaker spread, with females showing a slightly larger spread than men at 40Hz to 30Hz, respectively.

The QuakeBox speaker's show similar results to the MAONZE speakers with little within-speaker variation and a high between-speaker spread. However, like many of the parameters discussed above, the QuakeBox speakers as a whole show less uniformity than their MAONZE counterparts. The within-speaker variation in QuakeBox speakers is less consistent with only three speakers (QB02, QB03, and QB12) showing near equal antimode values. But like in the case of the MAONZE speakers the rest of the QuakeBox speakers don't fall too far behind, showing reliably low within-speaker variability across the entire speaker group. The most variable speakers, QB01 and QB16, show much more within-speaker variation than the rest of the QuakeBox speakers, but overall there is not a huge difference between each speaker's two antimode values. The male QuakeBox speakers' frequency spread is, like the MAONZE males, around 30. However, the majority of men stay within 20Hz. The female QuakeBox speakers once again show more spread than males, this time showing a between-speaker spread of over 100Hz. Two females (QB12 and QB15) show similar antimode values to the male speakers. This suggests that the antimode on one hand may be speaker specific, and on the other hand that it isn't necessarily as gendered as some other f0 descriptors.

The antimode is a new parameter that has not been previously studied. It is the only parameter that looks at what a speaker does not (possibly cannot) do when speaking, as it is a measurement of the least frequent f0 value used by a speaker. With an average within-speaker difference of 3.9Hz, it shows very little within-speaker variation. Out of all the parameters studied, the antimode is the most consistent in the within-speaker condition. The antimode also has a reasonably large between-speaker spread. This is more apparent in females rather than males, but the males do have a reasonable spread across both corpora. These conditions show the antimode to be a potentially very useful speaker discriminant.

4.7 Summary of results

Nolan's (1983) criteria for an effective speaker discriminant in forensic speaker comparison included high between-speaker variability and low within-speaker variability as the top requirements. f0 is already known to encompass the three criteria of availability, robustness in transmission, measurability, but due to its variable nature it is seen to be lacking within the first two. In this section I will evaluate the parameters investigated to determine their applicability as speaker discriminants within within-speaker variation and between-speaker variation.

Splitting creak and modal phonation into to separate distributions has yielded positive results for the mean f0 parameter. Both creak and modal f0 distributions show little within-speaker variation with the average f0 difference (appendix C, tables 12 and 13) of all speakers (excluding the outliers that are discussed in the next chapter) being 4.2Hz and 5.9Hz for creak mean f0 and modal mean f0 respectively. However, the mean is still a fluid measurement: the shape of a distribution can influence it greatly. The mean f0 also has a large between-speaker spread.

The mode f0 parameters show similarities to their mean counterparts, albeit with a slightly larger within speaker range. The average differences for the creak and modal mode f0 parameters are 9.3Hz and 10.2Hz respectively. This is nearly twice that of the mean f0 parameters for both creak mode and modal mode. While the within-speaker variability is higher, the mode f0 parameters do show a slightly larger between-speaker spread than the mean f0 parameters. And by just looking at the mode f0 visuals in figures 17 and 18 (section 4.2) there is still a substantial amount of independence in each speaker; the larger within-speaker variation in the mode f0, while not ideal or as definitive as the mean f0 values, can still potentially be a useful speaker discriminate.

There is one way the mode f0 parameters are superior to the mean f0 parameters. The mode is a stable measurement, meaning it doesn't move around the frequency distribution as the mean f0 can. This is potentially useful as when the shape of a distribution changes, even only a small change, the mean f0 will change also. But since the modal f0 is just the most frequently used frequency produced by the speaker, it can only change if another (usually neighbouring) frequency bin is used more. A good example of this is the modal f0 of speaker M05, which will be discussed in detail in the next chapter (section 5.3).

The skew and kurtosis did not appear to be effective speaker discriminants. While there was a large between-speaker spread for all skew and kurtosis measures, each measure was too variable within a speaker.

The most promising result was that of the antimode parameter. This measure has never been used before as a speaker discriminant, and was only found here by accident as a measure to separate each speaker's individual creak and modal distributions. The average antimode difference for all speakers is 3.9Hz, making it the best within-speaker measurement out of all the parameters investigated. It also shows large between-speaker variation, especially among females.

Overall the antimode parameter outshines the other parameters investigated in this study both within-speaker and between-speaker. The mode f0 parameters and the mean f0 parameters are similar in the way they behave both within-speaker and between-speaker, with the mean f0 parameters showing much more tight within-speaker variation. However, the mode provides stability as opposed to the mean's fluidity. The skew and kurtosis measurements did not show any effectiveness as speaker discriminants.

CHAPTER 5

Outliers

There are three male speakers from the MAONZE corpus who show inconsistent behavior throughout all measured parameters. These inconsistencies affect not only the results or within-speaker variation, but on between-speaker variation as well. Each speaker shows a different kind of irregularity to the standard bimodal distribution: M03 has a trimodal distribution, M06 has a unimodal distribution, and M05 has an idiosyncratic creak percentage. The three speakers and their distributional characteristics are discussed in detail below. Any mention of dual bimodal speaker averages within this chapter refer to tables 12 and 13 of appendix C, which show differences between each parameter for each individual speaker and their mean values.

5.1 M03 – Trimodal Distribution

As seen in figure 29, speaker M03 shows a trimodal frequency distribution in his Maori speech. There are three clear peaks: one near 50Hz (in his creak range), one near 100Hz (in his modal range) and one near 175Hz. This last peak shows a pitch frequency nearly equal to the creak distribution, and also shows a very wide positive skew that hasn't yet reached zero at the density plot's cut-off point. His English on the other hand shows a leptokurtic kurtosis with very little skew, indicating that his frequency range in English is much more limited than in Maori. His English also reaches an antimode nearing zero density around 150 Hz, though the bumpiness of the distribution line past this point indicates there were a few pitch measurements recorded in the higher pitch range of this speaker. This shows that speaker M03 spent a more substantial amount of time speaking within the higher frequency ranges while speaking Maori than in English.

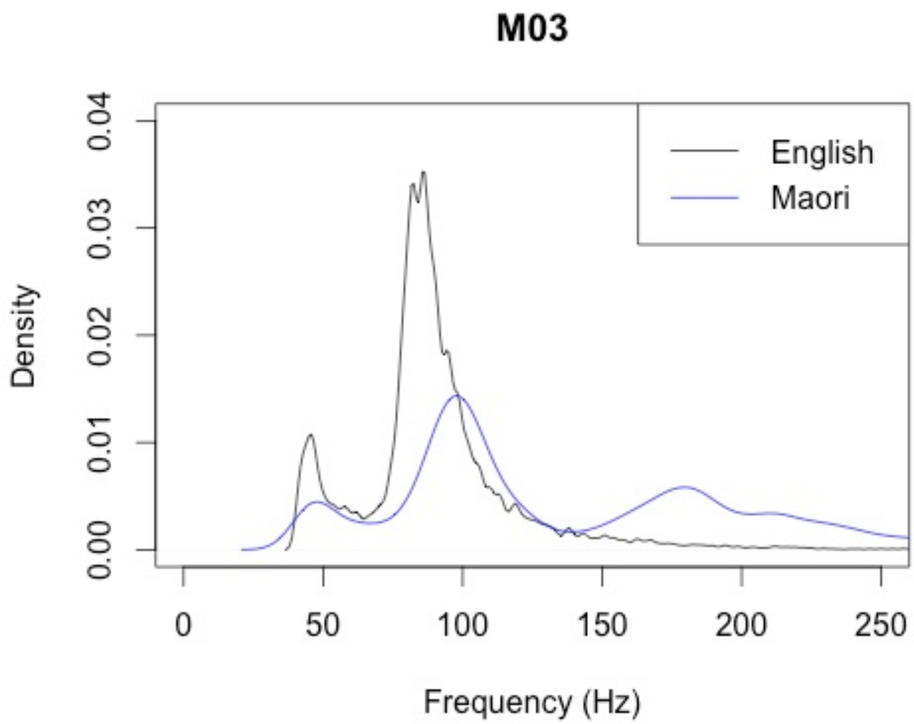


Figure 29 - Frequency distributions for M03

The different quantity of peaks in this speaker's two frequency distributions has sizeable effects on nearly all f0 parameters. As the difference mainly concerns the higher pitch ranges creak remains similar between both distributions with mode values of 46Hz (English) and 47Hz (Maori) and mean values of 49.5Hz and 50.8Hz. The skew and kurtosis of M03's creak is also similar between each distribution, with differences of only 0.06 and 0.1 respectively. M03's creak is relatively stable within the creak range, but in its modal and upper ranges they appear to separate considerably. This is because when calculating for bimodality anything past the antimode is counted as in the modal range, which in this case includes the third peak. Firstly the modal mean for M03's English is 98.6Hz while for Maori it is 144.8, a difference of 46.2Hz. This is substantially larger than the average man f0 for all dual bimodal speakers (speakers with two bimodal distributions) in this study, who have an average modal mean f0 difference of 5.9Hz. This massive difference of 46.2Hz is due to the fluidity of the mean along a distribution. The large amount of f0 measurements in the third peak pushes up the mean to a point in the distribution that would be near the second antimode (the point of lowest frequency between the second and third peaks). The modal mode is a much more stable measurement, however it still does show more difference between M03's distributions than a speaker with two bimodal distributions. M03's modal mode for English is 86Hz and 96Hz for Maori, with a difference of 10Hz. The average difference of the modal mode between all dual bimodal speakers in this study is 10.2Hz. M03 does fall within this range, showing that mode does not appear to be largely affected by a third peak in the upper frequency ranges. The skew and kurtosis of M03's modal distribution also vary. This can clearly be seen on the distribution plot above. His skew value is 2.4, slightly higher than the average 2.3. His kurtosis is shows a much more extreme difference at exactly 13, with the average among all dual bimodal speakers being 9.3. M03 is also one of the few speakers' who has a leptokurtic kurtosis in one distribution (his English) and a platykurtic kurtosis in the other (His Maori). As well as the dissimilarities in the skew and kurtosis in each distribution, it is unclear on what is exactly measured. This is because the R package used to calculate these values relies on the distribution to be unimodal, yet the modal distribution for M03's Maori speech is clearly bimodal. M03's antimode is also similar between his English and Maori, with a difference of 2.48Hz. This is well under the average dual bimodal speaker average of 3.9Hz.

Overall M03's Maori trimodal distribution has a considerable effect in the comparison of each f0 parameter in the modal range. Predictably his skew and kurtosis are inconsistent across his two languages, especially in his kurtosis. However, the modal measurements that show more consistency in most speakers are also not as effective in comparing M03's two languages. The modal mean is the shows the biggest difference at 46.2Hz, a 40.3Hz difference compared to the average modal mean of dual bimodal speakers. This is due to the fluid nature of the mean f0. There is a large amount of f0 measures above the modal peak, which pushes the modal mean higher. His modal mode measurement falls within the dual bimodal speaker average, with difference of only 0.2Hz between M03 and the collective dual bimodal speakers. Out of all the parameters the antimode performs the best for comparing both of M03's f0 distributions, with a difference of 2.48Hz, which is a closer comparison to the average dual bimodal speaker. The larger differences in M03's distributions can be attributed to the third peak in the upper frequency range. It is possible that these differences can be diminished if the upper pitch range was omitted from the data, leaving only two peaks. This would also correct any miscalculations of skew and kurtosis, as the modal distribution would be only one peak.

5.2 M06 – Unimodal Distribution

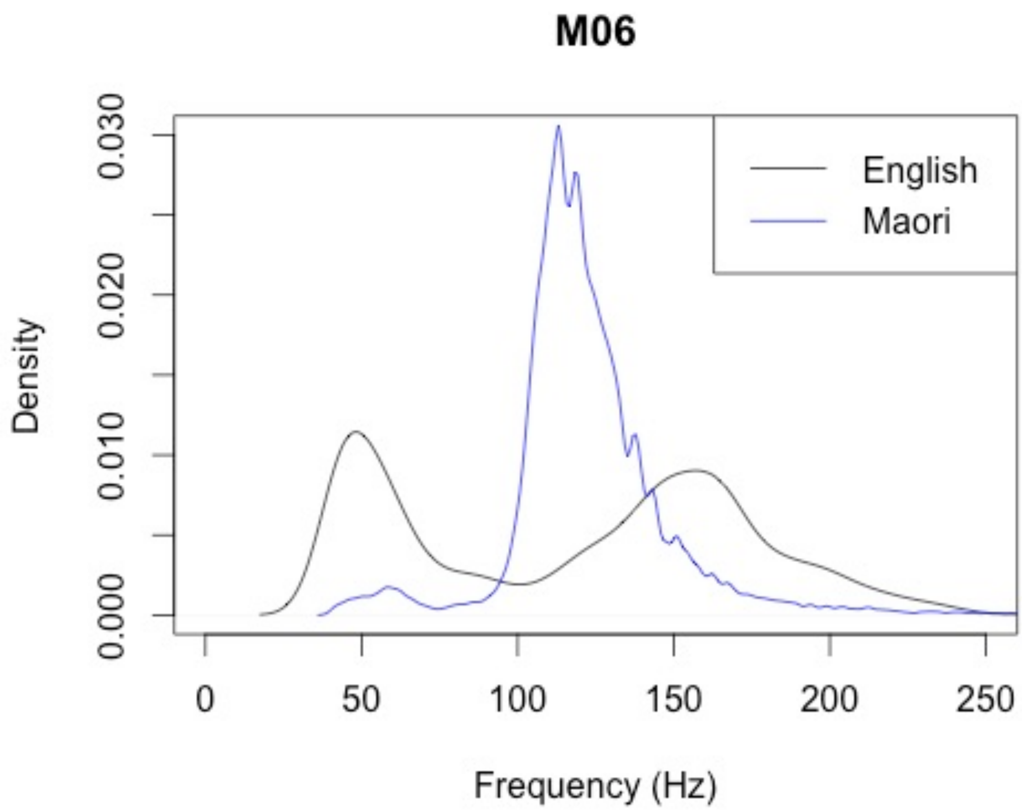


Figure 30 - Frequency distributions for M06

Figure 30 shows M06's English and Maori frequency distributions. His English is clearly bimodal, showing two peaks at around 50Hz and 150Hz. His Maori distribution, however, does not show two distinct peaks. There is a slight bump in his creak range, just above 50Hz. Only 3.9% of M06's total frequency measurements were recorded to be in the creak range (under the antimode), compared to the average dual bimodal speaker's 21.3%. This shows that creak is very infrequent in M06's Maori speech, leading to the conclusion that M06's Maori speech is unimodal rather than bimodal. The vast majority of M06's frequency measurements fall into the modal range, making M06 appear to be more unimodal than bimodal in his Maori frequency distribution. As there were some f0 measurements in the creak range an antimode was calculated and an analysis splitting the two distributions could be conducted.

The differences in M06's two distributions have large effects on the comparability of the two distributions. Firstly the antimode of each distribution is one of the only antimodes throughout all speakers in this study to show substantial difference between English and Maori, with the antimodes being 100.6Hz and 74Hz respectively. This large difference of 26.4Hz is a far cry from the dual bimodal speaker average of 3.9. The creak ranges are the least effected, even though one of the distributions has a small amount of creak. This isn't to say that the creak distributions aren't affected at all by M06's Maori unimodality; the creak modes for M06 are 41Hz (English) and 58Hz (Maori), a difference of 17Hz. On the other side, the creak means for M06 are 57Hz and 56.8Hz, a difference of 0.2Hz. The skew and kurtosis of the creak distributions show differences of 1.1 for skew and 0.6 for kurtosis. While not definitive, this does indicate that despite the large differences in the amount of creak M06 has in each distribution they are distributed similarly.

The modal distributions do not share the same relative closeness as the creak distributions do. This is obvious from just looking at the distribution plot in figure 30 above. The peak of the M06's English modal distribution is well outside his Maori modal distribution. This is also shown clearly in the values that represent these two distributions. The modal modes for M06 are 163Hz for English and 113Hz for Maori. This is a massive difference of 50Hz. The modal mean shows less difference than its mode counterpart with a 34.1Hz difference (160Hz for English and 125.9Hz for Maori). Ironically one of the best measurements for comparing these two distributions is the total mean, calculating the mean treating creak phonation and modal phonation as a single distribution. This gives mean values of 118Hz for English and 123Hz for Maori, a difference of 5Hz. This is not surprising as the peak of M06's Maori distribution is relatively close to the middle of his two English distributions. This however exemplifies some of the issues of calculating the total mean, as it is clear that the two distributions are very different. The skew and kurtosis of M06's modal distribution also don't show the similarities seen in the creak distribution. The difference in his skew is 2.4 and his kurtosis is very dissimilar with a difference of 23.8. Again this is obvious from just looking at the frequency distributions in Figure 30, they are clearly very different.

The issue with M06 is less about his unimodality and more about his two distributions being very different from each other. M03 had an extra peak affecting his mean measurements, but his other parameters such as mode and antimode were essentially unaffected. M06 does not show any similarities in his f0 parameters apart from the creak mean and total mean. There is very little creak in M06's Maori speech, which makes it hard to assess its validity as a speaker discriminant for this speaker. Also the total mean has been shown to be more inaccurate than other measures with speakers who have similar pitch distributions. The fact that it is one of the most similar parameters for this speaker is most probably a coincidence. It is obvious from just the distribution plot in figure 30 that M06 does not adhere to the norm when it comes to f0 distributions since his two distributions are very different from each other.

5.3 M05 – Extreme Distributions

M05 is another speaker that does not display a bimodal distribution in their two languages. Figures 31 and 32 show the frequency distributions for the English speech and Maori speech of speaker M05. Figure 31 is the full distribution while figure 32 is the zoomed in distribution as the extreme height of the full distribution obscures many details about both the English and Maori distributions. The most notable feature of M05 is his extremely frequent use of creak as seen in the huge peak in figure 31.

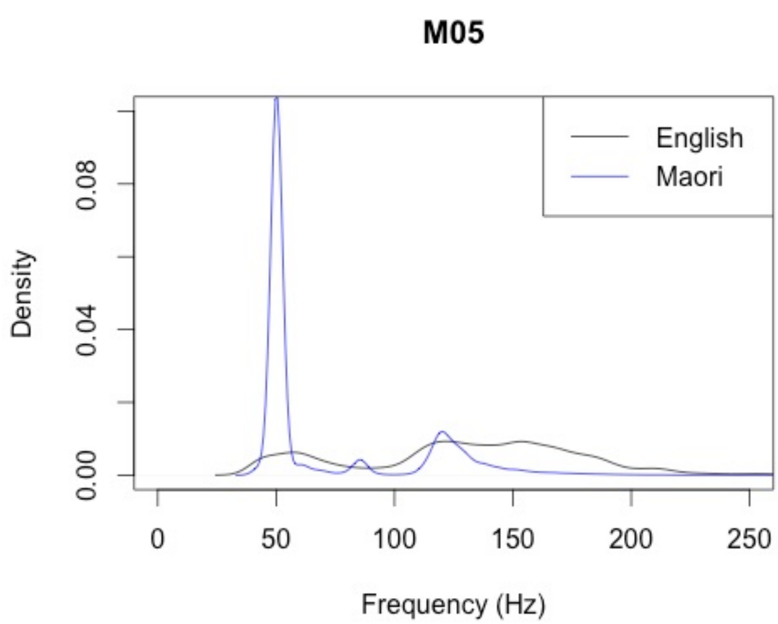


Figure 31 - frequency distributions for M05 - full

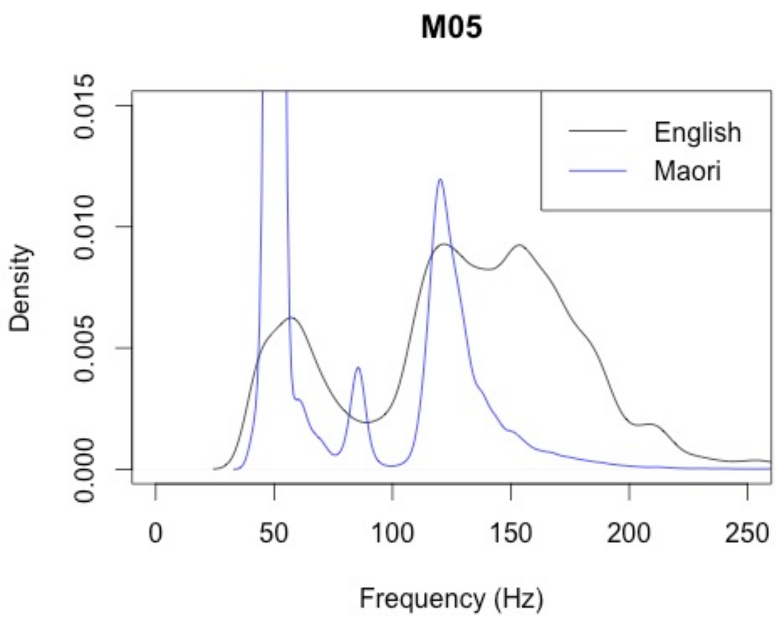


Figure 32 - Frequency distribution for M05 - zoomed in

M05 spends the majority of his time in creak, with 70.8% of his total measurements occurring under the antimode, as opposed to his English creak measurements, which sits at 20.5%. There also appears to be a small peak between the two main peaks, this could be considered a trimodal distribution. However, for the purposes of this study this small peak will be counted within creak phonation as it occurs to the left of the lowest antimode. M05 also shows a wider distribution in his English modal phonation, as opposed to a slightly taller, yet skinnier, peak in his Maori modal phonation.

These anomalies in M05's distributions affect many f0 parameters. Firstly his antimode, unlike the majority of other speakers, does not show clear similarities between English and Maori. His English antimode is 89Hz while his Maori antimode is 99.5Hz, a difference of 10.5Hz. This is above the average dual bimodal speaker, which sits at 3.9Hz. M05's creak modes for each language are 57Hz for English and 50 Hz for Maori, and his creak means are 59.3Hz and 50.7Hz respectively. M05's creak mode has a difference of 7Hz, which is under the average dual bimodal speaker of 9.3Hz. M05's creak mean difference (9.4Hz) does not fall under the dual bimodal speaker average, which sits at 4.2Hz. Once again creak skew and kurtosis don't show similarities, though this (especially kurtosis) is to be expected based on the distribution plot figure 32. M05's creak skew has a between language difference of 3.4 while his creak kurtosis has a difference of 20.4.

M05's modal values are also very dissimilar to each other. There is a lot more modal phonation in M05's English speech than in his Maori speech, and his English speech has a much wider phonation range. M05's modal means are 150.3Hz for English and 126Hz for Maori, with a difference of 23.7Hz. This is to be expected due to the larger amount of f0 measurements that occupy the upper range of his modal phonation. These f0 measurements push up the mean. The mode is not affected by these upper range f0 measurements, instead being found in the most frequently used measurement. For M05's modal phonation the modal mode is 119 for both English and Maori, achieving the best possible result of zero difference for speaker comparison. Predictably skew and kurtosis again do not provide any similarities between M05's languages, with skew showing a difference of 1.6 and kurtosis showing a difference of 15.5.

M05 has some irregular qualities that obscure potential speaker comparison. The huge amount of creak did not do too much to impair the creak results. That appears to be in part due to the small peak in between creak phonation and modal phonation. This however only appears to have effected the creak mean, which had a difference between languages higher than the average dual bimodal speakers difference. The same is true for M05's modal phonation, where the modal mean for both English and Maori were very different from each other. The opposite is true for the modal mode, which had the best possible result of zero difference. Overall despite the large amount of difference between M05's two f0 distributions the mode parameters showed similarities between them where other parameters couldn't.

5.4 Outlier Conclusions

The speaker's M03, M05, and M06 all show that f0 distributions are not always regular bimodal distributions. This can pose problems such as in the calculation of skew and kurtosis. These measurements, as unreliable as they seem to be in speaker comparison and discrimination, do rely on unimodal distributions to be able to be calculated accurately. M03 showed a trimodal distribution with a third peak in the upper pitch range. This is possibly easily fixable by cutting off the third peak at the second antimode, though this does remove potentially useful data. M06's distributions are not so easy to fix, as his two distributions were, through most parameters, completely different from each other. There does not appear to be an easy fix for M05 either, although while most parameters did not show similarities between the two distributions the modes for both creak and modal phonation did. The creak mode was closer in similarity than the dual bimodal speaker average, and the modal mode showed the best result possible with no difference between M05's English and Maori distribution. Not enough is known about the speakers in question to be able to identify reasons for this movement away from the norm of dual bimodal distributions.

While most speakers tend to behave similarly across distributions, these three speakers highlight that each speaker is highly unique. It is clear that while speakers generally have bimodality for both their distributions, it is not impossible for them to display unimodal or multimodal distributions. It is also possible for speakers to show hugely different distributions across their languages, as seen in speaker M05. The proposed method of separating creak phonation and modal phonation works only if both distributions from a speaker are bimodal. Those who deviate from this form of bilingual bimodality must be treated separately in order to take into account the differences in distributions. It may also be that speaker comparison based on f0 distributions is simply not possible for some speakers, such as M06.

CHAPTER 6

Conclusions

6.1 Evaluation of predictions

In section 2.7 various predictions were generated based on the current literature on f0 and forensic speaker comparison. This section will evaluate those predictions in the context of the current study.

Prediction 1: The use of the REAPER pitch tracker will show significant improvements on the accuracy of current pitch trackers.

It is well known throughout the literature that popular pitch trackers, such as Praat, are less than accurate in accounting for a speaker's full frequency distribution. In most cases Praat does not accurately observe the lower pitch ranges, which disconnects a speaker's creak distribution from the overall analysis. REAPER shows accuracy within the entire frequency range. While this study kept the pitch floor at the default 40Hz, REAPER has been shown to still be effective at tracking even lower frequencies.

As well as showing accuracy in the whole frequency range, REAPER is also very reliable across all speakers. Many studies using Praat or other pitch trackers had to manipulate individual speakers pitch range in order to gain an "accurate" measure of their pitch, as factors such as octave jumps distorted the original data.

Gold (2014) exemplifies this: many of her speakers needed their pitch boundaries to be altered in a hit-and-miss fashion in order to get rid of anomalies such as octave jumps. This, however, might lead to potentially inconsistent and inaccurate pitch measurements across speakers. A speaker with a pitch boundary of 75-220Hz for example will have a very different mean f0 than with a pitch boundary of 40-400Hz. REAPER, being a very accurate pitch tracker through the entire frequency range, does not need to be manipulated for each individual speaker to show accurate results, minimizing analyst bias in the work process.

Therefore, prediction 1 was supported. REAPER has been shown to be an accurate pitch tracker through the whole frequency range and provides more consistent results than other pitch trackers that have preceded it.

Prediction 2: There will be significant variation in f0 between an individual speaker's two languages.

Nearly all of the literature claims f0 to be language specific. However, most speakers in this study displayed similarities between their two languages. This was consistent across multiple parameters in both corpora and in all languages. The only speakers who showed consistent differences between languages in multiple parameters were those with irregular distributions. Speakers who showed bimodal distributions in both of their languages showed similar measurements in their creak modes and means, their modal modes and means, and their antimodes.

Therefore, based on the current study, prediction 2 was not supported. Since the majority of speakers show similar values across all parameters, f0 is probably better described as a speaker specific bundle. This may be because of the bilingual nature of the speakers. Pitch qualities may be transferred from a speaker's first language to their second language as they are learning it. This would account for the similarities across each speaker's languages. However, there is still a large amount of variability in speakers' f0 values across the MAONZE corpus, and also in the English distribution for speakers of both corpora.

Prediction 3: A homogenous group of bilingual speakers will show less variation in f0 than a heterogeneous group of bilingual speakers.

This prediction was based on the assumption that f0 is language specific. If f0 is language specific then similar f0 values are expected to be used across all speakers of that language. If prediction 3 were to be supported then the speakers of the MAONZE corpus should show less variation than the speakers of the QuakeBox corpus, as the MAONZE corpus has only one language pair (English and Maori) while the QuakeBox corpus has multiple. This does not appear to be the case. There is similar between-speaker spread across both corpora in nearly all f0 parameters. Males tend to have a smaller spread in both corpora, but overall there does not appear to be an adherence to a single f0 measurement in any f0 parameter.

Therefore prediction 3 is not supported by the current study. There is no substantial difference between the between-speaker spread of the MAONZE speakers and the QuakeBox speakers.

Prediction 4: Creak phonation alone will be less useful than modal phonation as a speaker discriminant, however its relationship to modal phonation may prove useful.

Creak phonation has always been considered to play a smaller role in a speaker's total f0 distribution, with the majority of studies claiming that creak is the same in all speakers regardless of gender or language group. Based on the current study this assumption is not supported. The amount of creak a speaker produces does not seemed to be systematically changing in any condition, each speaker shows a substantial amount of difference in the amount of creak they use independent from what language they speak or their gender. The range of the creak phonation, as measured by the antimode, is also relatively independent in each speaker.

Creak phonation does show little within-speaker variability, but due to its smaller range it also has less between-speaker spread. There does appear to be a small connection between the mean f0 values of creak phonation and modal phonation, however this does not appear to be consistent across all parameters, or even across all speakers. This indicates that creaky phonation is a largely independent phenomenon from modal phonation.

Prediction 4 is supported by the results of this study. While its small range does hinder the effectiveness of creak parameters, its independence from modal phonation does provide more information on a speaker, especially about the amount of within-speaker variation. This shows that quantifying the creak phonation range can be further pursued despite its limitations.

Prediction 5: The addition of further f0 distribution parameters (mode/skew/kurtosis) will improve the accuracy in quantifying the variability of both in the within-speaker and between-speaker condition for use in speaker discrimination.

Most studies calculate the mean f0 and its standard deviation and nothing else. This has been shown in multiple cases not to be the best speaker discriminant, yet it continues to be the most popular measurement for speaker comparison. The implementation of other f0 parameters is not new, with studies using parameters such as the mode, skew, and kurtosis (e.g. Kinoshita et al., 2005, Kinoshita and Ishihara, 2010). Very few of these studies take bimodal distribution into account. REAPER showed that most speakers have a clear bimodal distribution. This study separates the two distributions into creak phonation and modal phonation and calculates multiple f0 parameters for both.

This separation automatically improved the accuracy of mean measurements. Previously studies tended to include creak phonation in the total mean f0, this potentially shifted the mean downwards depending on how much creak each speaker produced and how many cycles the tracker accurately recovered. By taking creak phonation out the modal mean f0 the within-speaker variation narrowed considerably. The mode f0 values were also shown to be reasonably good speaker discriminants. Both creak mode and modal mode showed less similarities within-speaker than their mean f0 counterparts, however they are more stable measurements as they do not move along the frequency distribution based on the spread of frequency measurements. The mode f0 values also showed relatively good between-speaker spread.

The skew and kurtosis are parameters that have been discussed as potential speaker discriminants, however the results of this study show that both the skew and kurtosis are far too variable both within-speaker and between-speaker to be effective measurements for speaker discrimination. This is consistent among both creak phonation and modal phonation, and through all languages. The most fruitful outcome of the study was that of the discovery of the antimode. This parameter had not been previously studied, yet out of all the parameters considered in this study it showed the least amount of within-speaker variability with a reasonably large between-speaker spread.

Overall, prediction 5 was supported by the current study. The addition of other f0 parameters shows much more information about a speaker than using only mean f0 and the standard deviation around the mean. The separation of a speaker's creak and modal distributions also improves the accuracy of results. This allows for a much more refined speaker analysis.

6.2 Recommendations

Based on the results above, several recommendations can be made in regards to the methodologies of pitch tracking and the analysis of f0 in the context of forensic speaker comparison.

Recommendation 1: Use REAPER to track pitch instead of less accurate pitch trackers such as Praat.

There has been a variety of methods to cope with inaccurate pitch trackers, such as removing creak phonation, including creak phonation as part of the total f0 value, or manipulating the pitch boundaries in individual speakers to offset pitch tracker inaccuracies. This leads to a substantial amount of inconsistencies, not only within a study but also across studies. REAPER is incredibly accurate to the point that there is no need to manipulate the data to suit the specific study. Using REAPER will create consistency throughout f0 studies, allowing results from different studies to be comparable in a way that is not attainable now due to different methodologies (and largely undocumented Praat settings) being used in different studies.

Recommendation 2: Use more than just the mean f0 in the speaker comparison workflow.

The mean f0 is the standard frequency measure, though it has repeatedly been shown to be very variable within a speaker. The mean f0 has been shown to be a good speaker discriminant, however it is not as stable as other measurements and can be affected by factors such as a wide distribution. Other parameters (such as the mode and antimode) are not affected by this, and also show reasonably small within-speaker variability and large between-speaker spread. Using multiple f0 parameters to identify a speaker or discriminate between two separate recordings can strengthen the overall result and show correlations between recordings that were not visible with just one parameter. Therefore it is better to use multiple parameters when investigating a speaker's frequency distributions.

Recommendation 3: Take into account bimodality when working with f0 frequency distributions.

It is clear that the vast majority of speakers have a bimodal frequency distribution. Most studies treated these two distributions as one, combining them into a single f0 value. This, while providing potentially inaccurate f0 values, also completely ignores potential discriminating factors that could be used for speaker discrimination and comparison. An accurate pitch tracker such as REAPER has the ability to show the frequency distribution of the lower pitch ranges hat other pitch trackers (such as Praat) could not. This provides the opportunity to work with creak phonation as a separate entity to modal phonation, essentially doubling the potential speaker discriminants in a single f0 distribution. This study showed that separating a speaker's total f0 distribution by creak and modal phonation provides better overall results both within-speaker and between-speaker.

The addition of a creak distribution within a speaker's total distribution also provides another f0 parameter: the antimode. This is the point of lowest frequency between the creak and modal distributions and has been shown in this study to be an incredibly accurate speaker discriminant, as it has the least amount of within-speaker variability while also have a large between spread. Overall taking into account a speaker's bimodality allows for a more accurate view of the speaker's total f0 distribution.

Recommendation 4: Don't discount creak as speaker discriminant.

Creak has often been written off as a useful speaker discriminant, mostly due to practitioners' not being able to track it effectively. There is also the belief that creak is qualitatively the same in all speakers and it does not vary. The current study showed that this is not the case. The quantity and location of creak each speaker produced were incredibly variable, even within the speaker. This further supports the notion that creak is an independent phonation mode. All the parameters quantified in this study showed creak phonation to be a good within-speaker discriminant, showing little variation within each speaker's two languages. It is however slightly less useful in between-speaker analysis, because as creak phonation falls into a smaller range than modal phonation there is less room for a speaker to produce creak. This means that there is a tighter overall spread in creak phonation, but it is clear that all speakers show idiosyncratic behavior in this phonation. Creak might therefore show some discriminating features that can serve as useful discriminants.

6.3 Future Directions

This study was just a first step showing that with an accurate pitch tracker, and an observance of creak phonation, f0 can be a useful speaker discriminant. There are many statistical models that can be applied to the observations in this study, such as the use of probability functions such as likelihood ratios or expressing equal error rates (EER). These statistical approaches are the new incoming standard in the field of forensic (speech) science and they determine the effectiveness of a parameter (or parameters), showing whether or not a feature is statistically significant. The next step in the continuance of this study will be to calculate these probabilities, and to see if the results reach thresholds deemed necessary to be used in casework.

This study focused on the different-language / same-style condition. This was useful in showing that speakers do appear to behave similarly across languages when speaking in the same interview style. However,it is rare for a real world case in forensic speaker comparison to have two speech samples in the same style. A speaker's f0 can change depending on their volume, their health, the emotional state, as well as a myraid of other factors. Therefore the next step will be to apply REAPER on same-language different-style conditions to see if the results shown in this study are comparable. This will be especially interesting for the antimode parameter. It was shown to be the most effective within-speaker discriminant for the same style condition set in this study, but the antimode as an f0 parameter has never been studied before so it remains to be seen if it can match its success here with different style conditions.

There is also potential to look at the frequency distribution in more fine grained detail. REAPER's accuracy provides a massive amount of detail with a frequency distribution, as can be seen in figure 33.

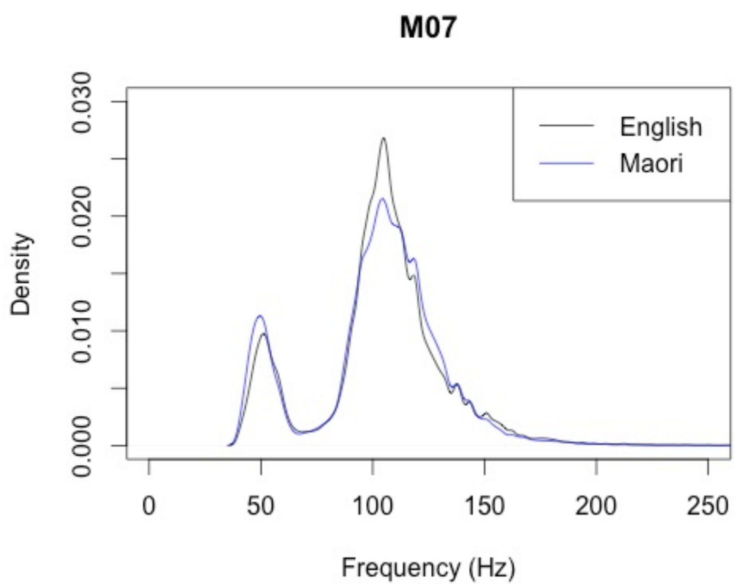


Figure 33 - M07 frequency distribution

The current study treated both the creak and modal distributions as single curves, however there are many bumps and crevices along M07's two distributions that may be very similar to each other. It may be possible to attain specific speaker information from a more fine-grained analysis of these frequency distributions.

Overall this study was just in many ways a starting point. The use of REAPER to accurately track pitch across a large group of speakers, the separation of creak phonation and modal phonation, and the use of the antimode to discriminate across speakers are all new concepts that require further study along different conditions.

APPENDIX A

Python code for removing interviewer and voicelessness

This code was created specifically for use in this thesis. Special thanks to Jesse Sheehan for producing it for me. This code can be downloaded from https://sheehan.nz/

```
#!/usr/bin/env python

"""A simple library to deal with TextGrid and f0 files."""

import re, sys, os

__author__ = 'Jesse Sheehan'
__copyright__ = 'Copyright 2017, Jesse Sheehan'
__license__ = 'GNU LGPLv3'
__version__ = '0.2'
__file__ = 'f0tool.py'
__email__ = 'jesse@sheehan.nz'

# Regular Expressions needed to parse the TextGrid files
_r_tg_xmin = re.compile('^[ \t]*xmin = (\d+\.\d+)$')
_r_tg_xmax = re.compile('^[ \t]*xmax = (\d+\.\d+)$')
_r_tg_all_tiers_begin = re.compile('^item \[\]\:$')

_r_tgt_begins = re.compile('^[ \t]*item \[(\d+)\]\:$')
_r_tgt_name = re.compile('^[ \t]*name = "([^"]*)"$')

_r_tgti_begins = re.compile('^[ \t]*intervals \[(\d+)\]\:$')
_r_tgti_text = re.compile('^[ \t]*text = "([^"]*)"$')

_r_f0_frame = re.compile('^(\d+\.\d+) (\d) (\-?\d+\.\d+)$')
_r_f0_frames_begin = re.compile('^EST_Header_End$')

# Regular Expressions needed to match certain file names
_r_f0 = re.compile('^(.+)\.f0$')
_r_tg = re.compile('^(.+)\.TextGrid$')

# The TextGrid Class
class TextGrid:
```

```
  xmin = 0.0
  xmax = 0.0
  tiers = []

  def __init__(self, fname):
    state = 'TextGrid'

    self.tiers = []
    self.xmin = 0.0
    self.xmax = 0.0

    f = open(fname)
    for line in f:

      # There are two regex patterns that are 'stateless' as they can occur whilst still in more
than one state

      # New Interval
      _ = _r_tgti_begins.match(line)
      if not _ is None:
        self.tiers[-1]['intervals'].append({'text': '', 'xmin': 0.0, 'xmax': 0.0})
        state = "Interval"
        continue

      # New Tier
      _ = _r_tgt_begins.match(line)
      if not _ is None:
        self.tiers.append({'name': '', 'xmin': 0.0, 'xmax': 0.0, 'intervals': []})
        state = "Tier"
        continue

      # Everything else has state:

      if state == 'TextGrid':
        # TextGrid xmin
        _ = _r_tg_xmin.match(line)
        if not _ is None:
          self.xmin = float(_.group(1))
          continue

        # TextGrid xmax
        _ = _r_tg_xmax.match(line)
        if not _ is None:
          self.xmax = float(_.group(1))
```

```
    continue

  # TextGrid tiers collection (item)
  _ = _r_tg_all_tiers_begin.match(line)
  if not _ is None:
    state = 'Tier'
    continue

elif state == 'Tier':

  # Tier xmin
  _ = _r_tg_xmin.match(line)
  if not _ is None:
    self.tiers[-1]['xmin'] = float(_.group(1))
    continue

  # Tier xmax
  _ = _r_tg_xmax.match(line)
  if not _ is None:
    self.tiers[-1]['xmax'] = float(_.group(1))
    continue

  # Tier name
  _ = _r_tgt_name.match(line)
  if not _ is None:
    self.tiers[-1]['name'] = _.group(1)
    continue

elif state == 'Interval':

  # Interval xmin
  _ = _r_tg_xmin.match(line)
  if not _ is None:
    self.tiers[-1]['intervals'][-1]['xmin'] = float(_.group(1))
    continue

  # Interval xmax
  _ = _r_tg_xmax.match(line)
  if not _ is None:
    self.tiers[-1]['intervals'][-1]['xmax'] = float(_.group(1))
    continue

  # Interval Text
  _ = _r_tgti_text.match(line)
```

```
      if not _ is None:
        self.tiers[-1]['intervals'][-1]['text'] = _.group(1)
        continue

    f.close()

  def __str__(self):
    t = ""
    for tier in self.tiers:
      t += "    " + tier['name'] + ": " + str(len(tier['intervals'])) + ' intervals\n'
    return "TextGrid Object\n  " + \
      "xmin: " + str(self.xmin) + "\n  " + \
      "xmax: " + str(self.xmax) + "\n  " + \
      "tiers:\n" + t

# The F0 class:
class F0:

  frames = []
  filename = ''

  def save(self, filename=None):
    if filename is None:
      filename = self.filename
    f = open(filename, 'w')

    f.write('EST_File Track\n')
    f.write('DataType ascii\n')
    f.write('NumFrames ' + str(len(self.frames)) + '\n')
    f.write('NumChannels 1\n')
    f.write('FrameShift 0.00000\n')
    f.write('VoicingEnabled true\n')
    f.write('EST_Header_End\n')

    for frame in self.frames:
      f.write('{:.6f}'.format(frame['time']) + ' ' + str(frame['voicing']) + ' ' +
'{:.6f}'.format(frame['pitch']) + '\n')

    f.close()

  def __init__(self, filename):
    self.filename = filename
    f = open(filename, 'r')
    state = 'Header'
```

```
    self.frames = []

    for line in f:
      if state == 'Header':
        _ = _r_f0_frames_begin.match(line)
        if not _ is None:
          state = 'Frames'
          continue
      elif state == 'Frames':

        self.frames.append({'time': 0.0, 'voicing': 0, 'pitch': 0.0})

        _ = _r_f0_frame.match(line)
        if _ is None:
          raise ValueError('error occurred reading f0 file :( line: "' + line + '"')
        else:
          self.frames[-1]['time'] = float(_.group(1))
          self.frames[-1]['voicing'] = int(_.group(2))
          self.frames[-1]['pitch'] = float(_.group(3))

    f.close()

  def __str__(self):
    return 'F0 Object - contains ' + str(len(self.frames)) + ' frames'

# Usage Functions:
def print_usage():
  print('USAGE: f0tool.py <Action>\n' + \
  '  Runs a particular action on an f0 file\n\n' + \
  '  Possible actions are:\n' + \
  '    RemoveSilence\tRemoves silence from an f0 file\n' + \
  '    IsolateTier\t\tIsolates a single tier on an f0 file\n' + \
  '    BulkRemoveSilence\tRemoves silence on a folder full of f0 files\n' + \
  '    BulkIsolateTier\tIsolates a single tier on a folder full of f0 files\n' + \
  '    BulkProcessData\tRuns BulkIsolateTier and BulkRemoveSilence on a folder')

def print_remove_silence_usage():
  print('USAGE: f0tool.py RemoveSilence <InputF0File> [OutputF0File]\n' + \
  '  Removes all frames that have no voicing from the f0 file\n\n' + \
  '  <InputF0File> is the f0 file to remove the silent frames from\n' + \
  '  <OutputF0File> is the file to be written. If ommited then <InputF0File> is used')

def print_isolate_tier_usage():
```

```
 print('USAGE: f0tool.py IsolateTier <TextGridFile> <TierNumber>|0 <InputF0File> [OutputF0File]\n' + \
 ' Remove all other tiers except for TierNumber from the specified f0 file.\n\n' + \
 ' <TextGridFile> specifies the file containing the tiers\n' + \
 ' <TierNumber> is the tier to isolate, if 0, you will be prompted\n' + \
 ' <InputF0File> is the file to read the frames from\n' + \
 ' <OutputF0File> is the file to write the frames to, if ommitted then <InputF0File> is used')

def print_bulk_remove_silence_usage():
 print('USAGE: f0tool.py BulkRemoveSilence <InFolder> [OutFolder]\n' + \
 ' Removes silence from all f0 files in a folder\n\n' + \
 ' <InFolder> is the folder containing the f0 files\n' + \
 ' [OutFolder] is where to write the resulting f0 files. If not specified, the original files are overwritten')

def print_bulk_isolate_tier_usage():
 print('USAGE: f0tool.py BulkIsolateTier <InFolder> <TierNumber>|0 [OutFolder]\n' + \
 ' Runs isolate_tier on a folder containing TextGrid and f0 files\n\n' + \
 ' <InFolder> is the folder containing the TextGrid and f0 files\n' + \
 ' <TierNumber> is the tier number to isolate. If 0 then you will be prompted\n' + \
 ' [OutFolder] is the folder where the f0 files should be written to. If ommitted, the original f0 files will be overwritten.')

def print_bulk_process_data_usage():
 print('USAGE: f0tool.py BulkProcessData <InFolder> <TierNumber>|0 [OutFolder]\n' + \
 ' Isolates the tiers then removes silence from all f0 files in a folder\n\n' + \
 ' <InFolder> is the folder containing the f0 files\n' + \
 ' <TierNumber> is the tier to isolate. If 0, you will be prompted\n' + \
 ' [OutFolder] is where to write the resulting f0 files. If not specified, the original files are overwritten')

def print_title():
 copyright = '| ' + __file__ + ' v' + __version__ + ', ' + __copyright__ + ' and licensed under ' + __license__ + ' |'
 bar = (len(copyright)-2) * '-'
 print(' ')
 print('\t/' + bar + '\\')
 print('\t' + copyright)
 print('\t\\' + bar + '/')
 print(' ')
```

```
# REMOVE SILENCE
def wrap_remove_silence(args):
  if len(args) > 1:
    infile = args[1]
    outfile = infile
    if len(args) > 2:
      outfile = args[2]
    remove_silence(infile, outfile)
  else:
    print_remove_silence_usage()

def remove_silence(infile, outfile):
  if not os.path.isfile(infile):
    print('ERROR: input file "' + infile + '" does not exist')
    sys.exit(1)

  sys.stdout.write('INFO: Removing all frames with silence from "' + infile + '" and
saving to "' + outfile + '"... ')
  sys.stdout.flush()

  f = F0(infile)
  new_frames = []

  for frame in f.frames:
    if frame['voicing'] == 1:
      new_frames.append(frame)
  f.frames = new_frames

  f.save(outfile)

  print('done')

# ISOLATE TIER
def prompt_for_tiernumber(tiers):
  print('PROMPT: Please select a tier:')
  i = 1
  for tier in tiers:
    print(' ' + str(i) + ') ' + tier['name'])
    i += 1
  line = -1
  while line < 1 or line > len(tiers):
    line = input('? ')
    try:
      line = int(line)
```

```
    except:
      line = -1
  return line

def isolate_tier(textgridfile, tiernumber, infile, outfile):
  if not os.path.isfile(textgridfile):
    print('ERROR: file "' + textgridfile + '" does not exist')
    sys.exit(1)

  if not os.path.isfile(infile):
    print('ERROR: file "' + infile + '" does not exist')
    sys.exit(1)

  if tiernumber < 0:
    print('ERROR: TierNumber must be 0 or positive')
    sys.exit(1)

  tg = TextGrid(textgridfile)

  if tiernumber > len(tg.tiers):
    print('ERROR: TierNumber must be less than ' + len(tg.tiers))
    sys.exit(1)

  if tiernumber == 0:
    tiernumber = prompt_for_tiernumber(tg.tiers)

  tier = tg.tiers[tiernumber - 1]

  f0 = F0(infile)

  sys.stdout.write('INFO: Isolating tier "' + tier['name'] + '" from "' + textgridfile + '" in "'
+ infile + '" and saving to "' + outfile + '"... ')
  sys.stdout.flush()

  new_frames = []

  for frame in f0.frames:
    for interval in tier['intervals']:
      if interval['text'] != '':
        if frame['time'] >= interval['xmin'] and frame['time'] < interval['xmax']:
          new_frames.append(frame)
          break
  f0.frames = new_frames
```

```
  f0.save(outfile)

  print('done')

def wrap_isolate_tier(args):
  if len(args) > 3:
    textgridfile = args[1]
    tiernumber = int(args[2])
    infile = args[3]
    outfile = infile
    if len(args) > 4:
      outfile = args[4]
    isolate_tier(textgridfile, tiernumber, infile, outfile)
  else:
    print_isolate_tier_usage()

# BULK REMOVE SILENCE
def bulk_remove_silence(infolder, outfolder):
  if not os.path.isdir(infolder):
    print('ERROR: directory "' + infolder + '" not found')
    sys.exit(1)

  if not os.path.isdir(outfolder):
    print('INFO: creating directory "' + outfolder + '"')
    os.mkdir(outfolder)

  for name in os.listdir(infolder):
    _ = _r_f0.match(name)
    if not _ is None:
      remove_silence(os.path.join(infolder, name), os.path.join(outfolder, name))

def wrap_bulk_remove_silence(args):
  if len(args) > 1:
    infolder = args[1]
    outfolder = infolder
    if len(args) > 2:
      outfolder = args[2]
    bulk_remove_silence(infolder, outfolder)
  else:
    print_bulk_remove_silence_usage()

# BULK ISOLATE TIER
def bulk_isolate_tiers(infolder, tiernumber, outfolder):
  if not os.path.isdir(infolder):
```

```
    print('ERROR: directory "' + infolder + '" not found')
    sys.exit(1)

  if not os.path.isdir(outfolder):
    print('INFO: creating directory "' + outfolder + '"')
    os.mkdir(outfolder)

  for name in os.listdir(infolder):
    _ = _r_f0.match(name)
    if not _ is None:
      tgname = os.path.join(infolder, _.group(1) + '.TextGrid')
      if os.path.exists(tgname) and os.path.isfile(tgname):
        isolate_tier(tgname, tiernumber, os.path.join(infolder, name), os.path.join(outfolder,
name))

def wrap_bulk_isolate_tier(args):
  if len(args) > 1:
    infolder = args[1]
    tiernumber = int(args[2])
    outfolder = infolder
    if len(args) > 3:
      outfolder = args[3]
    bulk_isolate_tiers(infolder, tiernumber, outfolder)
  else:
    print_bulk_isolate_tier_usage()

# BULK PROCESS DATA
def bulk_process_data(infolder, tiernumber, outfolder):
  if not os.path.isdir(infolder):
    print('ERROR: directory "' + infolder + '" not found')
    sys.exit(1)

  if not os.path.isdir(outfolder):
    print('INFO: creating directory "' + outfolder + '"')
    os.mkdir(outfolder)

  print('isolating tiers...')
  bulk_isolate_tiers(infolder, tiernumber, outfolder)
  print('removing silence...')
  bulk_remove_silence(outfolder, outfolder)

def wrap_bulk_process_data(args):
  if len(args) > 1:
    infolder = args[1]
```

```
    tiernumber = int(args[2])
    outfolder = infolder
    if len(args) > 3:
      outfolder = args[3]
    bulk_process_data(infolder, tiernumber, outfolder)
  else:
    print_bulk_process_data_usage()

if __name__ == '__main__':
  print_title()
  if len(sys.argv) > 1:
    if sys.argv[1] == 'RemoveSilence':
      wrap_remove_silence(sys.argv[1:])
    elif sys.argv[1] == 'BulkRemoveSilence':
      wrap_bulk_remove_silence(sys.argv[1:])
    elif sys.argv[1] == 'IsolateTier':
      wrap_isolate_tier(sys.argv[1:])
    elif sys.argv[1] == 'BulkIsolateTier':
      wrap_bulk_isolate_tier(sys.argv[1:])
    elif sys.argv[1] == 'BulkProcessData':
      wrap_bulk_process_data(sys.argv[1:])
    else:
      print_usage()
  else:
    print_usage()
```

APPENDIX B

R code used for all values and figures

```
# read file for both languages
lang1 <- read.csv("lang1.csv")
lang2 <- read.csv("lang2.csv")

#Create density plot for first language
plot (density(lang1$f0), main = "title", xlab = "Frequency (Hz)", xlim=c(0, 300),
ylim=c(0, 0.025), col = "black")

#add density plot for second language
lines(density(lang2$f0), col = "blue")

#add legend
legend("topright", c("English", "Maori"), col = c("black", "blue"), lty = 1)

#find antimodes
library("modes")
amps(lang1[,3])
amps(lang2[,3])

#split data into creak and modal
lang1_modal <- lang1[ which(lang1$f0 > antimode), ]
lang1_creak <- lang1[ which(lang1$f0 < antimode), ]
lang2_modal <- lang2[ which(lang2$f0 > antimode), ]
lang2)creak <- lang2[ which(lang2$f0 < antimode), ]

#find modes
amps(lang1_creak[,3])
amps(lang1_modal[,3])
amps(lang2_creak[,3])
amps(lang2_modal[,3])

#find skew
skewness(lang1_creak[,3])
skewness(lang1_modal[,3])
skewness(lang2_creak[,3])
skewness(lang2_modal[,3])

#find kurtosis
kurtosis(lang1_creak[,3])
kurtosis(lang1_modal[,3])
kurtosis(lang2_creak[,3])
```

```
kurtosis(lang2_modal[,3])

#find mean
mean(lang1_creak[,3])
mean(lang1_modal[,3])
mean(lang2_creak[,3])
mean(lang2_modal[,3])

#ggplot bar graph
ggplot(lang1_creak, aes(x=subject.language, y=creak.parameter, fill=subject, group=subject)) + geom_bar(stat="identity") + geom_line(stat="identity") + theme(axis.text.x=element_text(angle=45,hjust=1,vjust=1)) + ggtitle("title") + theme(plot.title = element_text(lineheight=.8, face="bold"))
```

APPENDIX C

Tables of figures and parameter values

Table 3 - MAONZE f0 measurements amounts and creak %

MAONZE f0 measurements amounts and creak %					
speaker	language	total f0 measures	# modal measures	# creak measures	creak %
M01	English	165052	136066	28986	17.56173812
	Maori	182886	142752	40134	21.94481808
M02	English	90520	66314	24206	26.7410517
	Maori	103909	72081	31828	30.63064797
M03	English	134176	98448	35728	26.62771285
	Maori	39076	33721	5355	13.70406388
M04	English	67418	28400	39018	57.87475155
	Maori	231304	174304	57000	24.6428942
M05	English	26048	20717	5331	20.46606265
	Maori	559000	163360	395640	70.7763864
M06	English	13757	8944	4813	34.9858254
	Maori	148280	142514	5766	3.888589156
M07	English	154685	129597	25088	16.21876717
	Maori	140078	114233	25845	18.45043476
M08	English	111795	65039	46756	41.82297956
	Maori	145220	108944	36276	24.9800303
M09	English	186587	172914	13673	7.327948892
	Maori	130540	122399	8141	6.236402635
M10	English	322412	209825	112587	34.92022629
	Maori	122460	89127	33333	27.21950024
M11	English	232836	205475	27361	11.75118968
	Maori	41734	35671	6063	14.5277232
M12	English	88711	69907	18804	21.19692034
	Maori	184679	154081	30598	16.56820754
M13	English	213619	177196	36423	17.05044963
	Maori	173511	142680	30831	17.76890226
M14	English	47908	39629	8279	17.28103866
	Maori	121146	99815	21331	17.60767999
M15	English	49262	41517	7745	15.72205757
	Maori	249790	211542	38248	15.31206213
M16	English	59693	48611	11082	18.56499087
	Maori	260815	218745	42070	16.13020724
M17	English	28552	22623	5929	20.76562062

Table 4 - MAONZE Modes and antimodes

MAONZE Modes and antimodes				
speaker	language	Antimode	modal mode	creak mode
M01	English	83.2832	119	58
	Maori	90.91701	119	62
M02	English	70.68209	119	47
	Maori	73.42744	94	47
M03	English	64.53732	86	46
	Maori	67.01564	96	47
M04	English	96.90664	119	42
	Maori	93.22223	137	52
M05	English	89.08802	119	43
	Maori	99.52931	119	50
M06	English	100.5854	163	41
	Maori	74.0392	113	58
M07	English	68.13979	105	52
	Maori	67.1375	103	50
M08	English	76.19239	101	50
	Maori	78.94579	101	51
M09	English	88.44008	119	62
	Maori	87.862426	119	62
M10	English	114.4096	155	45
	Maori	114.7735	178	47
M11	English	108.75177	167	58
	Maori	109.69362	178	58
M12	English	138.22301	184	50
	Maori	140.68026	190	56
M13	English	122.09344	184	68
	Maori	124.98527	188	54
M14	English	134.33696	188	62
	Maori	128.58342	168	62
M15	English	135.15453	186	62
	Maori	128.37484	172	62
M16	English	106.32295	158	66
	Maori	108.24753	154	50
M17	English	125.25576	174	54
	Maori	125.42283	193	58

Table 5 - MAONZE Means and standard deviations

MAONZE Means and standard deviations						
speaker	language	total mean	modal mean	modal sd	creak mean	creak sd
M01	English	120.4012	131.8361	30.33534	58.86286	8.802582
	Maori	123.3593	141.9339	34.75402	60.87822	9.521493
M02	English	97.69046	111.7303	24.05915	51.98374	8.019952
	Maori	92.1279	108.5657	20.80041	51.89072	7.772475
M03	English	90.1279	96.8397	26.05397	49.54651	6.611314
	Maori	134.6252	144.833	54.8497	50.78164	7.394902
M04	English	108.8539	148.5245	35.50073	53.75157	13.58542
	Maori	138.6267	151.4544	33.05977	62.79642	14.54339
M05	English	129.9628	150.272	32.83167	59.26539	12.56564
	Maori	72.92923	126.0447	26.62234	50.72536	3.29677
M06	English	118.8554	159.9712	30.2646	56.95768	15.1913
	Maori	123.2932	125.8134	25.82002	56.7804	8.188539
M07	English	103.7032	112.5183	21.39732	52.34805	5.928411
	Maori	102.0882	112.3706	19.87342	50.87261	5.640205
M08	English	84.97967	107.6958	20.68484	52.45849	7.099934
	Maori	93.94575	107.2833	17.21648	53.64266	7.873079
M09	English	134.9111	141.0115	22.72065	59.60317	9.143521
	Maori	137.8632	143.2881	23.26641	58.78122	9.432936
M10	English	135.1973	170.9196	29.35265	67.04551	17.69574
	Maori	149.2832	180.8445	30.95888	62.32362	19.09027
M11	English	164.0275	174.8484	28.31594	75.65732	17.10601
	Maori	169.9435	185.1981	34.85975	71.08738	19.45121
M12	English	166.4693	191.0519	27.78065	80.13777	23.7831
	Maori	177.1218	196.9223	26.36602	82.27134	23.98431
M13	English	177.5373	198.0287	46.49932	77.78709	19.24821
	Maori	177.5236	199.2498	41.04493	77.71824	20.75121
M14	English	173.1386	193.8852	33.36574	80.7286	24.029
	Maori	163.02	180.8954	29.87583	81.68292	20.8087
M15	English	177.0928	196.0002	36.48653	81.13407	23.5406
	Maori	170.9386	188.6349	33.13678	74.42577	21.21415
M16	English	155.706	169.5153	32.13527	76.5735	17.48379
	Maori	151.8855	164.1688	32.75179	74.68784	17.08936
M17	English	168.2164	193.2808	34.63676	73.33758	20.17301
	Maori	185.1423	204.1995	39.36037	75.96281	20.86079

Table 6 - MAONZE Skew and Kurtosis

MAONZE Skew and Kurtosis					
speaker	language	modal skew	creak skew	modal kurtosis	creak kurtosis
M01	English	2.375032	0.1833995	9.806966	0.2454377
	Maori	2.681191	0.1965735	11.68471	0.5556256
M02	English	1.649114	0.4726016	8.720213	-0.8915868
	Maori	1.729362	0.7865449	7.619914	-0.2000078
M03	English	3.031637	0.6072043	13.71047	-0.7872063
	Maori	0.6387719	0.6676204	-0.6931913	-0.6877116
M04	English	1.230632	1.204357	1.368748	0.8073164
	Maori	1.884708	0.3615079	6.473662	-1.06983
M05	English	0.838713	0.4269684	1.338684	-0.6605699
	Maori	2.406825	3.794589	16.81247	19.71008
M06	English	0.768864	1.079553	1.128241	0.241181
	Maori	3.57355	-0.06431589	24.89929	-0.7837731
M07	English	1.990552	0.3251975	8.94518	-0.2685345
	Maori	1.370149	0.4637408	4.901919	-0.2094927
M08	English	3.126464	0.8921413	12.98022	1.05905
	Maori	3.849515	0.95285	22.15563	0.9058134
M09	English	1.312554	-0.2175732	2.907223	0.3035992
	Maori	0.7378814	0.2620455	0.4861662	0.2548368
M10	English	1.589722	0.33562	4.732857	-0.7562872
	Maori	0.9825757	0.878112	2.407609	-0.367902
M11	English	2.648319	-0.322649	12.99395	-0.9775318
	Maori	1.950445	0.1369158	8.350413	-1.282106
M12	English	1.47757	0.1563329	3.782442	-0.854697
	Maori	1.634015	0	5.23174	-0.8959182
M13	English	1.596994	0.04333117	3.467531	-0.6029562
	Maori	1.295508	0.04932111	2.230916	-0.8434177
M14	English	1.818808	0.2007427	5.500726	-0.8297238
	Maori	1.599492	-0.02303068	5.1655	-0.732891
M15	English	2.502635	0.0989355	9.331049	-0.8934882
	Maori	3.172186	0.200248	16.45985	-0.9485759
M16	English	1.322847	-0.2141709	3.805583	-0.9450615
	Maori	2.53525	-0.05405932	10.4896	-0.790958
M17	English	1.733364	0.1334939	4.982153	-0.8146774
	Maori	2.282502	0.09060448	9.323346	-0.885166

Table 7 MAONZE f0 measurements amounts and creak %

MAONZE f0 measurements amounts and creak %					
speaker	language	total f0 measures	# modal measures	# creak measures	creak %
M01	English	165052	136066	28986	17.56173812
	Maori	182886	142752	40134	21.94481808
M02	English	90520	66314	24206	26.7410517
	Maori	103909	72081	31828	30.63064797
M03	English	134176	98448	35728	26.62771285
	Maori	39076	33721	5355	13.70406388
M04	English	67418	28400	39018	57.87475155
	Maori	231304	174304	57000	24.6428942
M05	English	26048	20717	5331	20.46606265
	Maori	559000	163360	395640	70.7763864
M06	English	13757	8944	4813	34.9858254
	Maori	148280	142514	5766	3.888589156
M07	English	154685	129597	25088	16.21876717
	Maori	140078	114233	25845	18.45043476
M08	English	111795	65039	46756	41.82297956
	Maori	145220	108944	36276	24.9800303
M09	English	186587	172914	13673	7.327948892
	Maori	130540	122399	8141	6.236402635
M10	English	322412	209825	112587	34.92022629
	Maori	122460	89127	33333	27.21950024
M11	English	232836	205475	27361	11.75118968
	Maori	41734	35671	6063	14.5277232
M12	English	88711	69907	18804	21.19692034
	Maori	184679	154081	30598	16.56820754
M13	English	213619	177196	36423	17.05044963
	Maori	173511	142680	30831	17.76890226
M14	English	47908	39629	8279	17.28103866
	Maori	121146	99815	21331	17.60767999
M15	English	49262	41517	7745	15.72205757
	Maori	249790	211542	38248	15.31206213
M16	English	59693	48611	11082	18.56499087
	Maori	260815	218745	42070	16.13020724
M17	English	28552	22623	5929	20.76562062
	Maori	27134	23130	4004	14.75639419

Table 8 QuakeBox total f0 measurements and creak %

QuakeBox total f0 measurements and creak %					
subject	langauge	total pitch measurements	#modal pitch measurements	#creak measurements	percentage of creak
QB01	english	61112	40891	20221	33.0884278
	french	66713	60053	6660	9.983061772
QB02	english	78762	58626	20136	25.56562809
	japanese	89627	47516	42111	46.98472558
QB03	english	180640	143756	36884	20.41851196
	mandarin	169570	144119	25451	15.00914077
QB04	english	92221	76599	15622	16.93974257
	mandarin	109721	81399	28322	25.81274323
QB05	english	25834	23702	2132	8.252690253
	mandarin	63076	44884	18192	28.84139768
QB06	english	120846	102602	18244	15.09690019
	maori	200824	120808	80016	39.84384337
QB07	english	60371	51880	8491	14.06469994
	maori	142048	82928	59120	41.61973417
QB08	english	45397	41121	4276	9.419124612
	punjabi	75481	52459	23022	30.50039083
QB09	english	35703	32930	2773	7.766854326
	russian	49294	39283	10011	20.30875969
QB10	english	33333	29689	3644	10.93210932
	cantonese	44295	31817	12478	28.17022237
QB11	english	160750	137263	23487	14.61088647
	german	164413	120344	44069	26.80384155
QB12	english	27473	25861	1612	5.867579078
	japanese	39143	34131	5012	12.80433283
QB13	english	30440	21329	9111	29.93101183
	japanese	55039	31173	23866	43.36197969
QB14	english	116483	95425	21058	18.0781745
	mandarin	218858	162913	55945	25.5622367
QB15	english	66163	40584	25579	38.66058069
	maori	46044	33448	12596	27.35644166
QB16	english	53730	39586	14144	26.32421366
	maori	93986	40024	53962	57.41493414
QB17	english	121275	102762	18513	15.26530612
	maori	183745	113450	70295	38.25682331

Table 9 - QuakeBox modes and antimodes

QuakeBox modes and antimodes				
subject	langauge	antimode	modal mode	creak mode
QB01	english	97.63909	147	45
	french	81.8459	125	51
QB02	english	86.59532	115	54
	japanese	85.02995	116	53
QB03	english	83.38108	132	49
	mandarin	85.13145	132	50
QB04	english	104.71055	125	68
	mandarin	100.88274	136	48
QB05	english	97.10667	125	57
	mandarin	93.39688	123	48
QB06	english	86.8793	111	59
	maori	93.02715	116	47
QB07	english	84.18276	125	59
	maori	91.35015	136	50
QB08	english	86.7888	144	68
	punjabi	90.08877	147	50
QB09	english	66.86306	86	43
	russian	69.49984	90	46
QB10	english	103.94753	155	78
	cantonese	108.4527	147	48
QB11	english	138.82957	181	86
	german	144.9566	198	49
QB12	english	82.3145	225	41
	japanese	82.6293	225	49
QB13	english	100.9482	179	50
	japanese	104.2058	147	54
QB14	english	124.61831	186	82
	mandarin	126.26264	195	50
QB15	english	76.00106	192	50
	maori	73.44079	183	50
QB16	english	118.49593	164	80
	maori	130.9559	168	49
QB17	english	109.32192	164	50
	maori	112.5443	147	43

Table 10 - QuakeBox means and standard deviations

QuakeBox means and standard deviations							
subject	langauge	total mean	total sd	modal mean	modal sd	creak mean	creak sd
QB01	english	118.4253	48.32904	148.456	25.60894	57.69688	14.86683
	french	131.0307	36.65566	139.023	28.98294	58.96425	10.73356
QB02	english	110.9085	39.53379	129.7862	26.12455	55.94608	8.235174
	japanese	94.52656	43.5631	130.0986	28.77302	54.38875	8.128048
QB03	english	120.9977	44.95139	138.257	32.48122	53.72942	9.900036
	mandarin	130.0104	44.4395	143.4499	33.18597	53.90793	10.34912
QB04	english	142.3348	35.25714	150.6098	25.46133	65.94046	16.3047
	mandarin	128.3062	48.22769	152.1742	29.21743	59.70808	14.63541
QB05	english	126.1466	31.12655	131.6217	25.91404	65.27832	15.36412
	mandarin	110.9999	42.17062	132.3766	29.01633	58.25835	13.40012
QB06	english	107.5659	25.8437	115.9414	17.14147	60.46303	12.58578
	maori	95.28604	35.29479	121.8208	14.3521	55.22387	12.14451
QB07	english	129.181	42.20015	140.7965	33.15376	58.20995	9.197636
	maori	105.2834	48.8334	141.5387	29.04767	54.4276	10.7733
QB08	english	140.9485	28.95242	148.9981	14.99684	63.5378	10.99383
	punjabi	118.8384	45.12387	147.6321	12.74604	53.2276	10.69104
QB09	english	89.51299	18.04224	92.76999	14.55535	50.83541	7.304717
	russian	90.17882	26.80988	100.3116	19.62366	50.41803	6.655458
QB10	english	169.2269	53.15007	181.4004	42.19975	70.04534	16.94438
	cantonese	143.6559	57.77131	174.4901	34.26053	65.28821	17.45054
QB11	english	175.4754	44.26749	190.9786	23.78447	84.87114	22.18305
	german	171.0443	64.65237	205.82	30.6627	76.07865	26.57008
QB12	english	205.6171	63.91286	214.6323	54.27626	60.98714	11.59068
	japanese	193.8662	73.85536	214.14	55.02093	55.80477	11.12747
QB13	english	143.5326	63.7546	178.9231	38.79876	60.68286	16.11212
	japanese	126.6871	69.25939	179.4488	43.56024	57.77146	14.08341
QB14	english	177.0936	60.07741	198.2131	42.84682	81.38981	21.56911
	mandarin	176.1536	74.44923	211.8791	47.3354	72.12049	24.87466
QB15	english	126.0425	70.72529	173.2718	48.40238	51.10772	8.177801
	maori	140.3029	69.85064	173.7664	51.00596	51.44228	7.545555
QB16	english	145.8284	50.00713	171.1112	28.39391	75.06721	21.14174
	maori	111.0401	63.16386	178.9709	27.86545	60.65537	20.28723
QB17	english	155.3416	48.08071	170.8203	33.11005	69.4225	18.6167
	maori	130.0536	62.40128	173.794	33.43241	59.46051	17.41449

Table 11 - QuakeBox skew and kurtosis

QuakeBox skew and kurtosis						
subject	langauge	total mean	skew of mode	skew of creak	kurtosis of mode	kurtosis of creak
QB01	english	118.4253	1.13301	1.00133	3.941874	0
	french	131.0307	1.216455	0.3052437	2.904096	-0.8609442
QB02	english	110.9085	2.495239	0.9943455	10.08401	1.713519
	japanese	94.52656	1.870675	1.06376	7.381717	1.601651
QB03	english	120.9977	2.285517	0.9320098	9.370109	0.09062651
	mandarin	130.0104	1.663969	0.944005	4.134508	0.1663798
QB04	english	142.3348	1.683193	0.3372422	4.824836	-0.7427622
	mandarin	128.3062	1.341531	0.7472321	2.312317	-0.2831063
QB05	english	126.1466	3.801624	0.4488686	23.16958	-0.8443714
	mandarin	110.9999	3.117442	0.7237541	13.32583	-0.5084443
QB06	english	107.5659	4.725661	0.3871289	42.74421	-0.8920062
	maori	95.28604	2.784966	1.053175	17.5773	0.34940802
QB07	english	129.181	2.026161	0.4285259	6.316554	0.1548041
	maori	105.2834	1.84143	1.161069	5.918653	0.8754324
QB08	english	140.9485	1.760708	-0.665629	6.083753	-0.7506641
	punjabi	118.8384	0.1062975	1.298539	7.307213	0.8747965
QB09	english	89.51299	4.23901	0.3889678	49.59829	-0.9453071
	russian	90.17882	4.05717	0.6391582	34.16273	-0.2231726
QB10	english	169.2269	2.239746	-0.03162705	9.066876	-1.114384
	cantonese	143.6559	1.427667	0.4303662	4.915645	-0.8783984
QB11	english	175.4754	2.525795	0.1616253	13.20292	-0.5043117
	german	171.0443	1.878108	0.5749452	6.763093	-0.6582944
QB12	english	205.6171	0.04060636	-0.2182316	1.785063	-1.050519
	japanese	193.8662	0.1150681	0.6819974	1.260693	-0.6907568
QB13	english	143.5326	0.7794664	0.7289685	1.293218	-0.602843
	japanese	126.6871	1.186573	1.156076	2.003712	0.7126015
QB14	english	177.0936	1.50346	0.09100695	3.214561	-0.8905638
	mandarin	176.1536	1.064459	0.524983	1.709626	-1.021639
QB15	english	126.0425	-0.4420286	1.162197	-0.1286519	0.8838471
	maori	140.3029	-0.4551613	0.8822029	0.0055205	0.1706991
QB16	english	145.8284	2.921029	0.1429983	17.86978	-1.066699
	maori	111.0401	3.229282	1.362189	19.57326	1.050421
QB17	english	155.3416	1.854306	0.2874504	6.742283	-0.9595561
	maori	130.0536	1.948554	1.148629	8.192271	0.5125853

Table 12 - differences between values

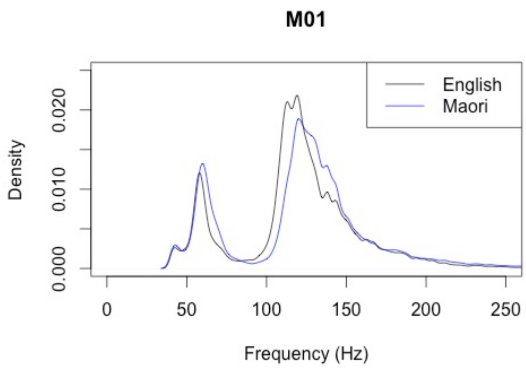
differences between values					
speaker	modal mode	creak mode	total mean	modal mean	creak mean
M01	0	4	25.66884	30.2036	2.01536
M02	25	0	5.56256	3.1646	0.09302
M04	18	10	29.7728	2.9299	9.04485
M07	2	2	1.615	0.1477	1.47544
M08	0	1	8.96608	0.4125	1.18417
M09	0	0	2.9521	2.2766	0.82195
M10	23	2	14.0859	9.9249	4.72189
M11	11	0	5.916	10.3497	4.56994
M12	6	6	10.6525	5.8704	2.13357
M13	4	14	0.0137	1.2211	0.06885
M14	20	0	10.1186	12.9898	0.95432
M15	14	0	6.1542	7.3653	6.7083
M16	4	16	3.8205	5.3465	1.88566
M17	19	4	16.9259	10.9187	2.62523
QB01	22	6	12.6054	9.433	1.26737
QB02	16	1	16.38194	0.3124	1.55733
QB03	0	1	9.0127	5.1929	0.17851
QB04	11	20	14.0286	1.5644	6.23238
QB05	2	9	15.1467	0.7549	7.01997
QB06	5	12	12.27986	5.8794	5.23916
QB07	11	9	23.8976	0.7422	3.78235
QB08	3	18	22.1101	1.366	10.3102
QB09	4	3	0.66583	7.54161	0.41738
QB10	8	30	25.571	6.9103	4.75713
QB11	17	37	4.4311	14.8414	8.79249
QB12	0	8	11.7509	0.4923	5.18237
QB13	32	4	16.8455	0.5257	2.9114
QB14	9	32	0.94	13.666	9.26932
QB15	9	0	14.2604	0.4946	0.33456
QB16	4	31	34.7883	7.8597	14.41184
QB17	17	7	25.288	2.9737	9.96199
MEAN	10.1935483	9.25806451	12.9751164	5.92489709	4.19123548

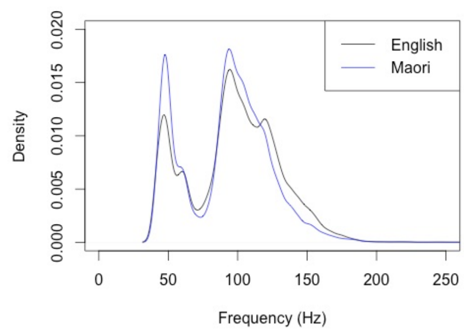
Table 13 - differences between values

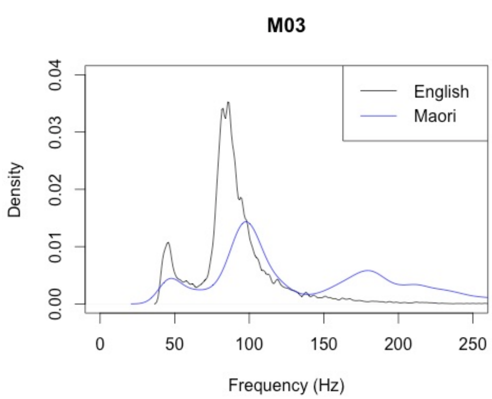
differences between values					
speaker	antimodes	modal skew	creak skew	modal kurtosis	creak kurtosis
M01	7.63381	0.306159	0.013174	1.877744	0.3101879
M02	2.74535	0.080248	0.3139433	1.100299	0.691579
M04	3.68441	0.654076	0.8428491	5.104914	1.8771464
M07	1.00229	0.620403	0.1385433	4.043261	0.0590418
M08	2.7534	0.723051	0.0607087	9.17541	0.1532366
M09	0.577654	0.5746726	0.4796187	2.4210568	0.0487624
M10	0.3639	0.6071463	0.542492	2.325248	0.3883852
M11	0.94185	0.697874	0.4595648	4.643537	0.3045742
M12	2.45725	0.156445	0.1563329	1.449298	0.0412212
M13	2.89183	0.301486	0.00598994	1.236615	0.2404615
M14	5.75354	0.219316	0.22377338	0.335226	0.0968328
M15	6.77969	0.669551	0.1013125	7.128801	0.0550877
M16	1.92458	1.212403	0.16011158	6.684017	0.1541035
M17	0.16707	0.549138	0.04288942	4.341193	0.0704886
QB01	15.79319	0.083445	0.6960863	1.037778	0.8609442
QB02	1.56537	0.624564	0.1317502	2.702293	0.111868
QB03	1.75037	0.621548	0.0119952	5.235601	0.07575329
QB04	3.82781	0.341662	0.4099899	2.512519	0.4596559
QB05	3.70979	0.684182	0.2748855	9.84375	0.3359271
QB06	6.14785	1.940695	0.6660461	25.16691	1.24141422
QB07	7.16739	0.184731	0.7325431	0.397901	0.7206283
QB08	3.29997	1.6544105	1.964168	1.22346	1.6254606
QB09	2.63678	0.18184	0.2501904	15.43556	0.7221345
QB10	4.50517	0.812079	0.46199325	4.151231	0.2359856
QB11	6.12703	0.647687	0.4133199	6.439827	0.1539827
QB12	0.3148	0.07446174	0.900229	0.52437	0.3597622
QB13	3.2576	0.4071066	0.4271075	0.710494	1.3154445
QB14	1.64433	0.439001	0.43397605	1.504935	0.1310752
QB15	2.56027	0.0131327	0.2799941	0.13417241	0.713148
QB16	12.45997	0.308253	1.2191907	1.70348	2.11712
QB17	3.22238	0.094248	0.8611786	1.449988	1.4721414
MEAN	3.86021593	0.53177469	0.44115959	4.259383523	0.553017887

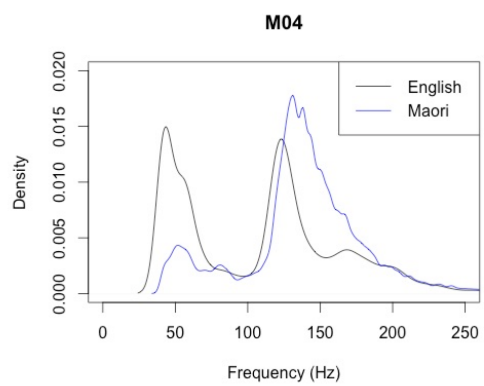
APPENDIX D

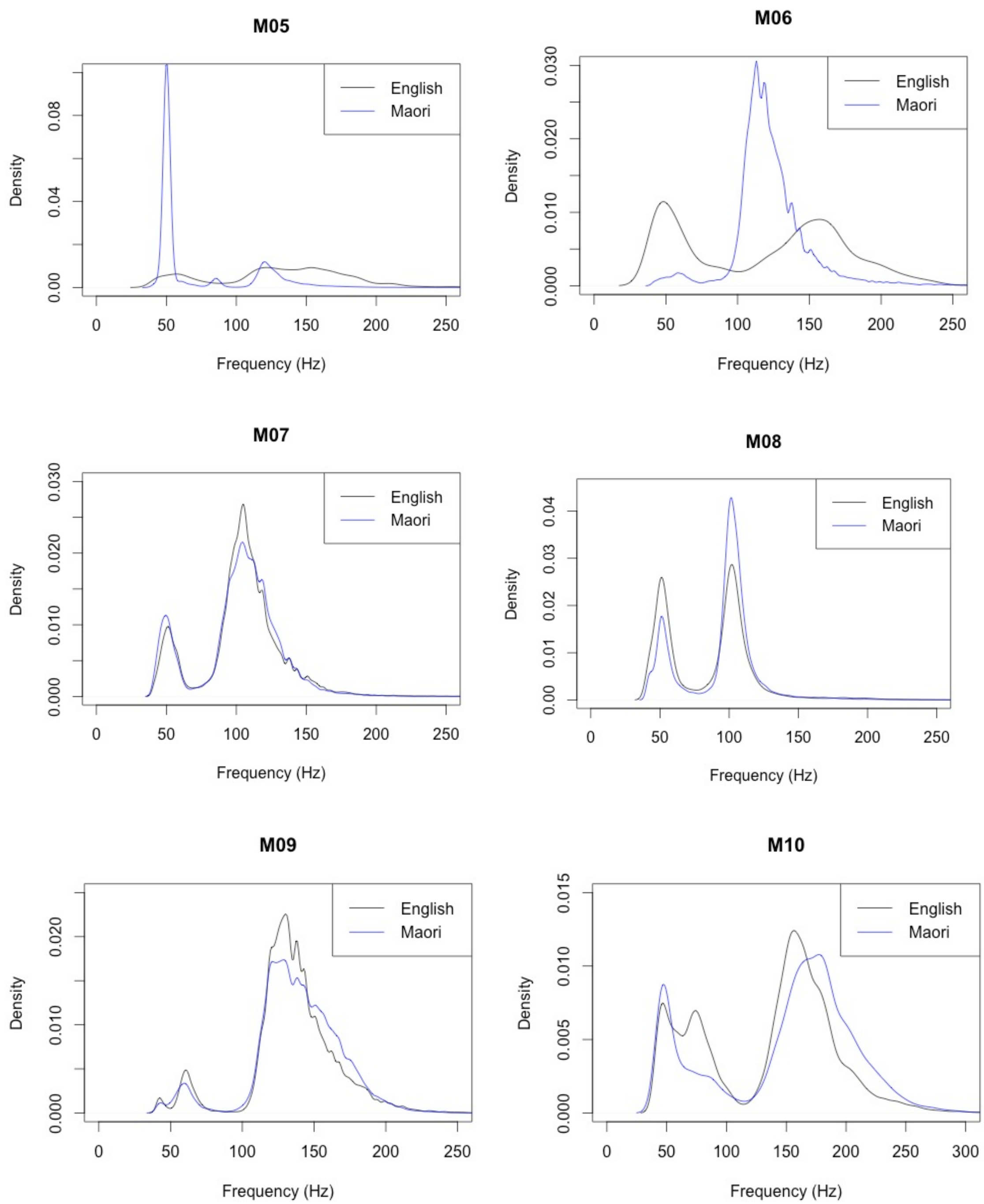
Density plots for all speakers

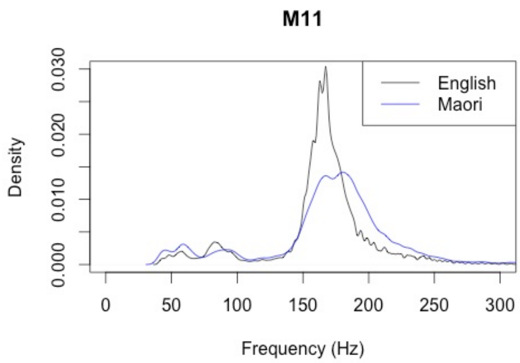


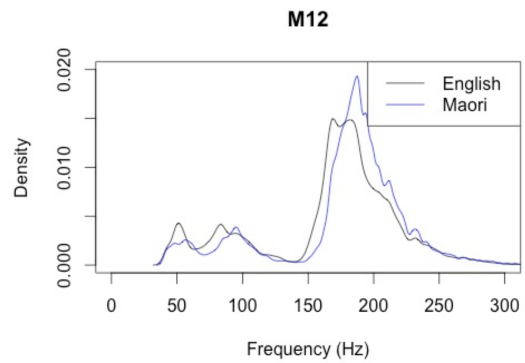


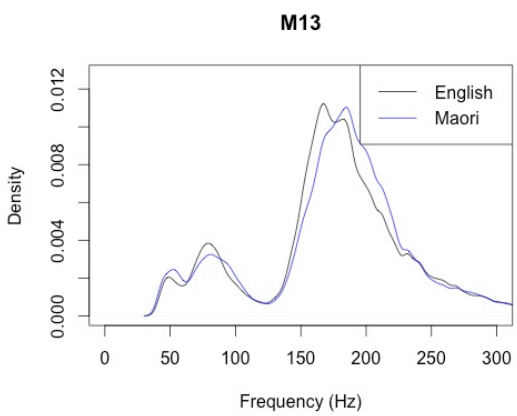


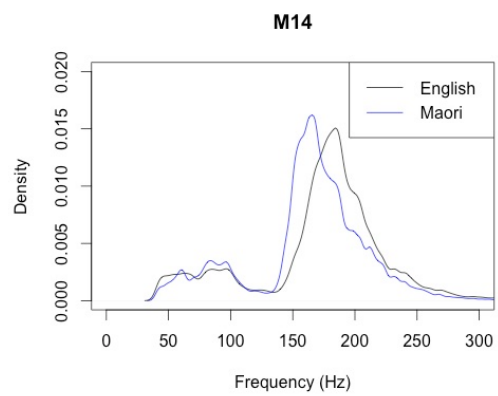


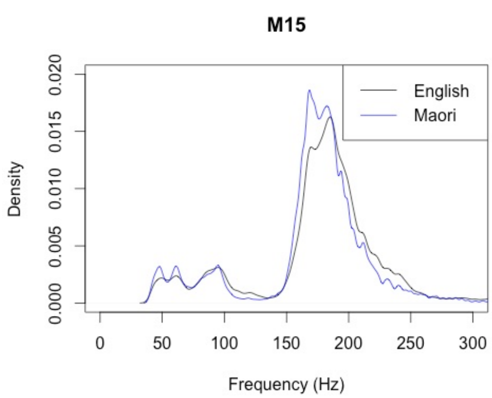

M05
M06
M07
M08
M09
M10
English
Maori
Density
Frequency (Hz)

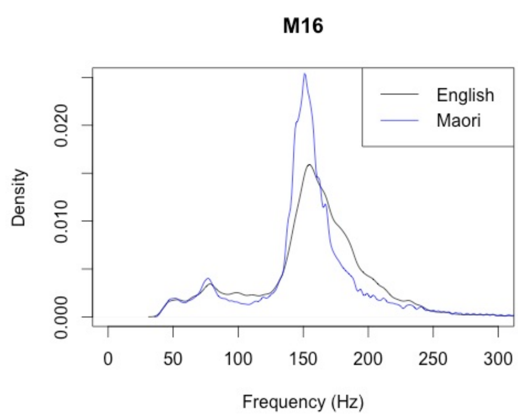

M11
English
Maori
Density
0.000
0.010
0.020
0.030
0
50
100
150
200
250
300
Frequency (Hz)

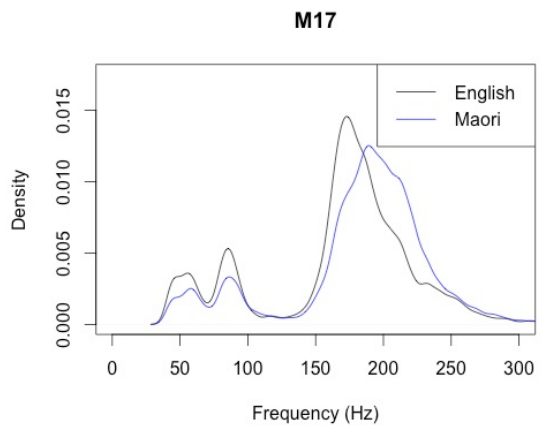

M12
English
Maori
Density
0.000
0.010
0.020
0
50
100
150
200
250
300
Frequency (Hz)

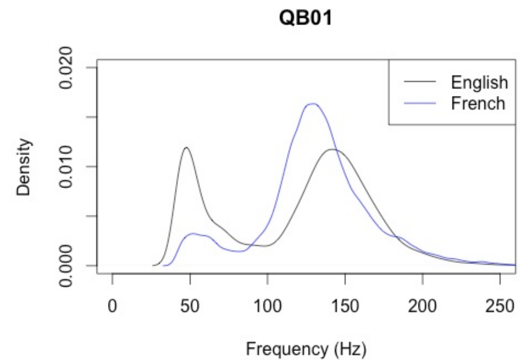

M13
English
Maori
Density
0.000
0.004
0.008
0.012
0
50
100
150
200
250
300
Frequency (Hz)

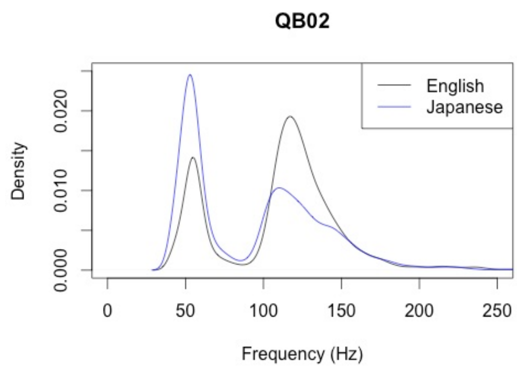

M14
English
Maori
Density
0.000
0.005
0.010
0.015
0.020
0
50
100
150
200
250
300
Frequency (Hz)

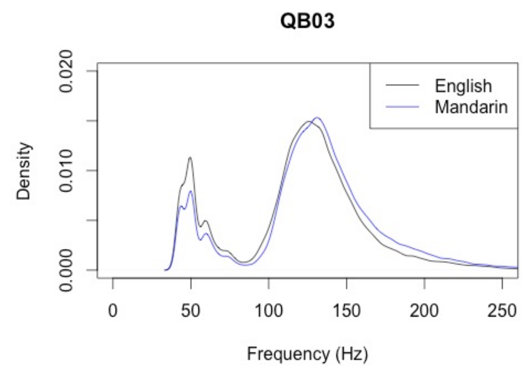

M15
English
Maori
Density
0.000
0.005
0.010
0.015
0.020
0
50
100
150
200
250
300
Frequency (Hz)

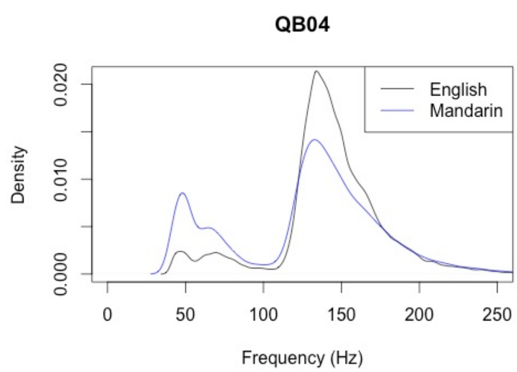

M16
English
Maori
Density
0.000
0.010
0.020
0
50
100
150
200
250
300
Frequency (Hz)

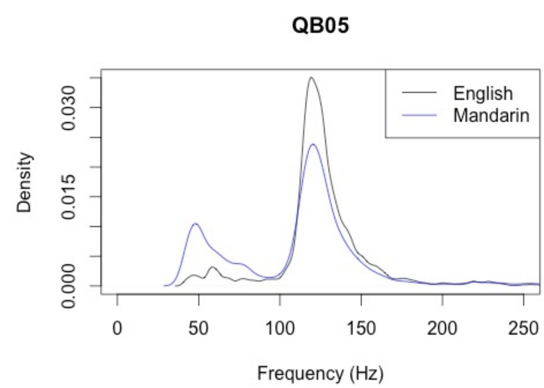

M17
English
Maori
Density
0.000
0.005
0.010
0.015
0
50
100
150
200
250
300
Frequency (Hz)

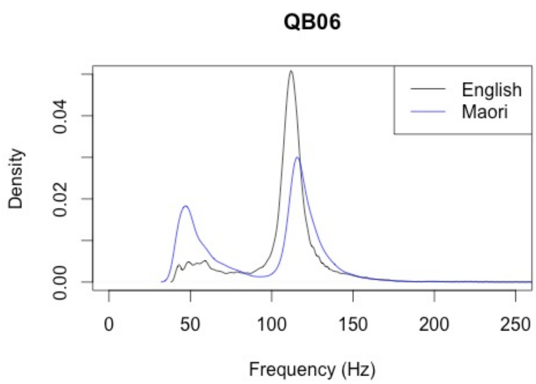

QB01
English
French
Density
0.000
0.010
0.020
0
50
100
150
200
250
Frequency (Hz)

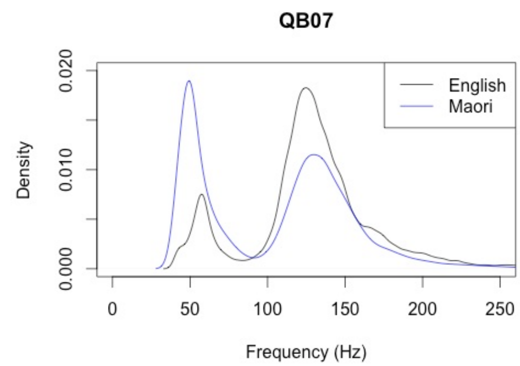

QB02
English
Japanese
Density
0.000
0.010
0.020
0
50
100
150
200
250
Frequency (Hz)

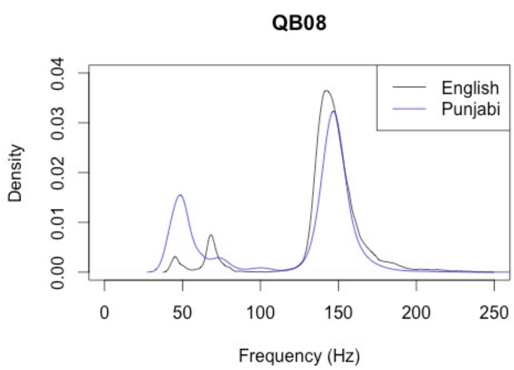

QB03
English
Mandarin
Density
0.000
0.010
0.020
0
50
100
150
200
250
Frequency (Hz)

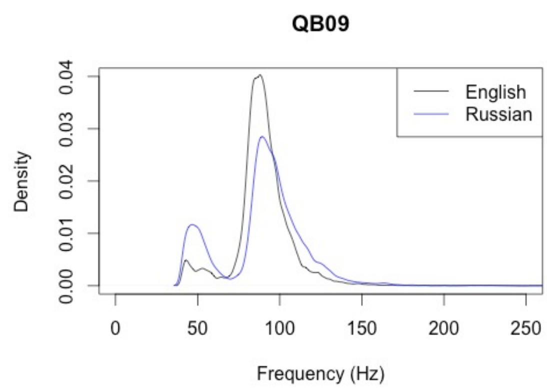

QB04
English
Mandarin
Density
0.000
0.010
0.020
0
50
100
150
200
250
Frequency (Hz)

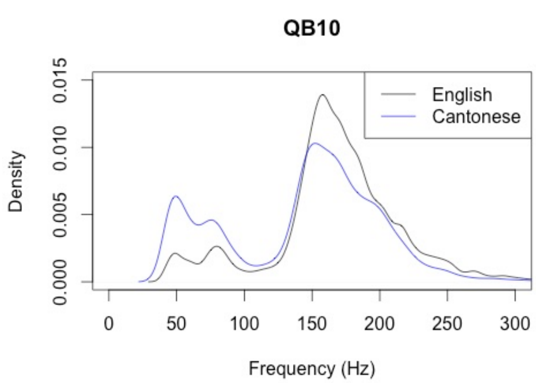

QB05
English
Mandarin
Density
0.000
0.015
0.030
0
50
100
150
200
250
Frequency (Hz)

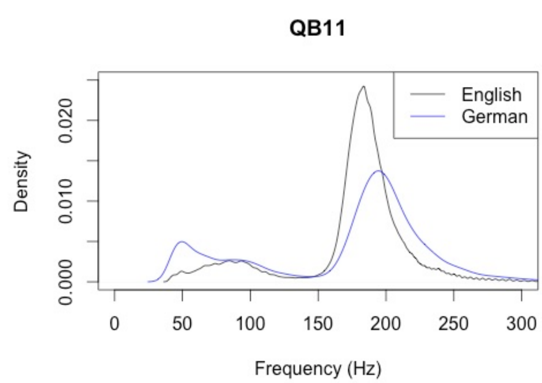

QB06
English
Maori
Density
0.00
0.02
0.04
0
50
100
150
200
250
Frequency (Hz)

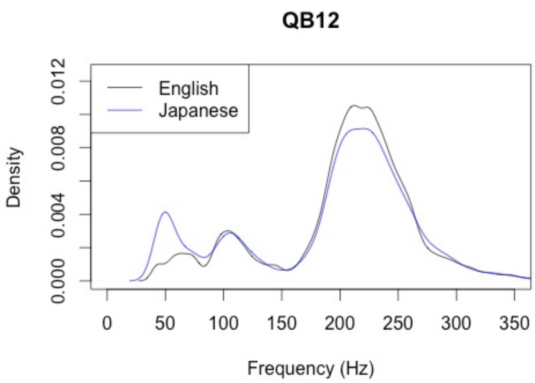

QB07
English
Maori
Density
0.000
0.010
0.020
0
50
100
150
200
250
Frequency (Hz)

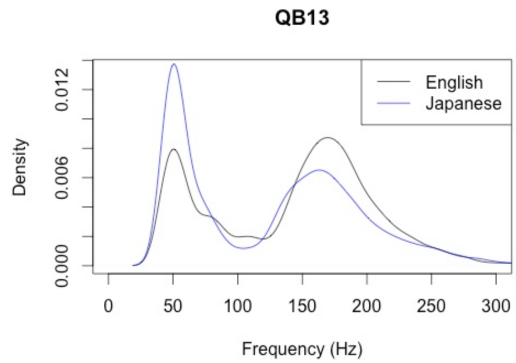

QB08
English
Punjabi
Density
0.00
0.01
0.02
0.03
0.04
0
50
100
150
200
250
Frequency (Hz)

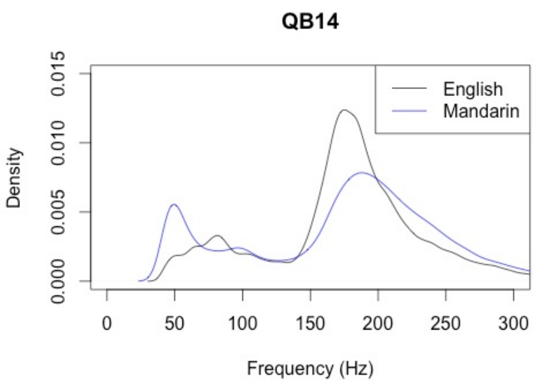

QB09
English
Russian
Density
0.00
0.01
0.02
0.03
0.04
0
50
100
150
200
250
Frequency (Hz)

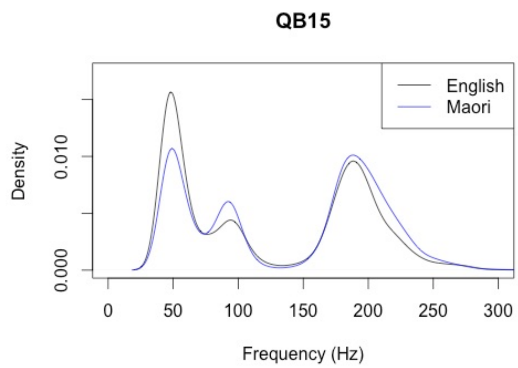

QB10
English
Cantonese
Density
0.000
0.005
0.010
0.015
0
50
100
150
200
250
300
Frequency (Hz)

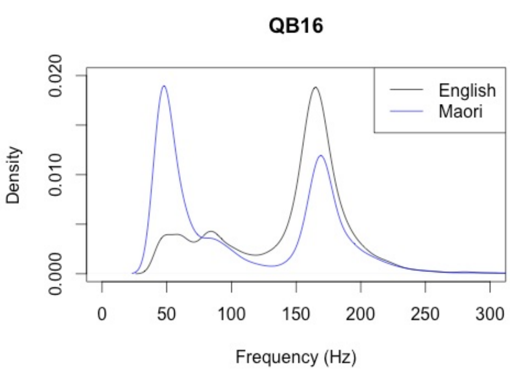

QB11
English
German
Density
0.000
0.010
0.020
0
50
100
150
200
250
300
Frequency (Hz)

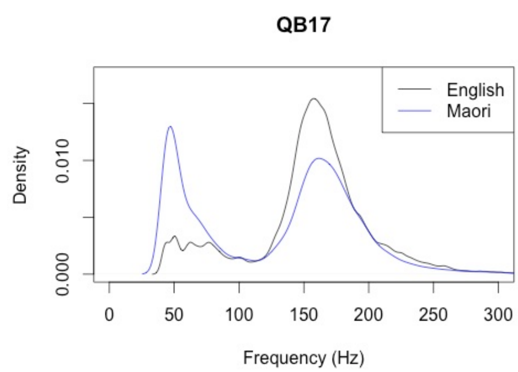

QB12
English
Japanese
Density
0.000
0.004
0.008
0.012
0
50
100
150
200
250
300
350
Frequency (Hz)


QB13
English
Japanese
Density
0.000
0.006
0.012
0
50
100
150
200
250
300
Frequency (Hz)


QB14
English
Mandarin
Density
0.000
0.005
0.010
0.015
0
50
100
150
200
250
300
Frequency (Hz)


QB15
English
Maori
Density
0.000
0.010
0
50
100
150
200
250
300
Frequency (Hz)


QB16
English
Maori
Density
0.000
0.010
0.020
0
50
100
150
200
250
300
Frequency (Hz)


QB17
English
Maori
Density
0.000
0.010
0
50
100
150
200
250
300
Frequency (Hz)

REFERENCES

Aare, K., Lippus, P., & Šimko, J. (2014). Creaky voice in spontaneous spoken Estonian. *XXVIII FONETIIKAN PÄIVÄT*, 27.

Abdelli-Beruh, N. B., Wolk L. & Slavin D. (2014). Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice, 28*(2), 185-190.

Ackerman, L., Hesterberg, L., & Bradlow, A. R. (2011). Talker and language variation in the F0 of English, Mandarin and Mandarin-accented English.

Boss, D. (1996). The problem of F0 and real-life speaker identification: A case study. *International Journal of Speech Langauge and the Law 3*(1), 155-159.

Bowen, L. K., Hands, G. L., Pradhan, S., & Stepp, C. E. (2013). Effects of Parkinson's disease on fundamental frequency variability in running speech. *Journal of medical speech-language pathology*, *21*(3), 235.

Clark, L., MacGougan, H., Hay, J., & Walsh, L. (2016). "Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus. *Ampersand*, *3*, 13-20.

Connell, B. (2002). Tone languages and the universality of intrinsic F 0: evidence from Africa. *Journal of Phonetics*, *30*(1), 101-129.

Deevi, S. and 4D Strategies (2016). modes: Find the Modes and Assess the Modality of Complex and Mixture Distributions, Especially with Big Datasets. R package version 0.7.0. http://CRAN.R-project.org/package=modes

Elliott, J. (2000). Comparing the acoustic properties of normal and shouted speech: a study in forensic phonetics. In *Proc. SST-2000: 8th Int. Conf. Speech Sci. & Tech* (pp. 154-159).

R. Fromont & J. Hay. (2008) ONZE Miner: the development of a browser-based research tool. *Corpora, 3*(2), 173-193.

Gold, E. (2014). *Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters* (unpublished doctoral thesis). University of York, York, U.K.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383-406.

Hudson, T., de Jong, G., McDougall, K., Harrison, P., and Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. In 16th Proceedings of the International Congress of Phonetic Sciences, Saarbrücken, pp. 1809-1812.

Jarvinen K., Laukkanen A-M., & Aaltonen O. (2013). Speaking a foreign language and its effect on the F0. Logopedics Phoniatrics Vocology, 38(2), 47-51.

King, J., Maclagan, M, Harlow, R., Keegan, P., and Watson, C. (2010). The MAONZE Corpus: Establishing a Corpus of Maori Speech, *New Zealand Studies in Applied Linguistics, 16*(2), 1-16.

Kinoshita Y. (2005). Does Lindley's LR estimation formula work for speech data? Investigation using long-term F0. International Journal of Speech, Language and the Law, 12, pp. 235-254.

Kinoshita, Y. and Ishihara, S. (2012). The effect of sample size on the performance of likelihood ratio-based forensic voice comparison. Proceedings of the 14th Australasian International Conference on Speech Science and Technology. Macquarie University, Australia.

Kinoshita, Y., & Ishihara, S. (2014). Background population: how does it affect LR-based forensic voice comparison?. *The International Journal of Speech, Language and the Law*, *21*(2), 191-224.

Kinoshita, Y., Ishihara, S., and Rose, P. (2009). Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. International Journal of Speech, Language and the Law, 16, pp. 91-111.

Künzel, H. J. (2000) Effects of voice disguise in speaking fundamental frequency. *International Journal of Speech Langauge and the Law, 7*(2), 150-179

Leemann, A., H. Mixdorff, M. O'Reilly, M.-J. Kolly, V. Dellwo (2014). Speaker-individuality in Fujisaki model f0 features: implications for forensic voice comparison. *International Journal of Speech, Language and the Law, (21)*.2, 343-370.

Liberman, M. (2015, February 8). REAPER. Retrieved from http://languagelog.ldc.upenn.edu/nll/?p=17590

Lindh J. (2006) Preliminary Descriptive F0-statistics for Young Male Speakers. *Lund University Working Papers*, *52*, 89-92.

Loakes, D. (2006). *A Forensic Phonetic Investigation into the Speech Patterns of Identical and Non-Identical Twins.* (Unpublished Ph.D. thesis), Melbourne University, Melbourne, Australia.

Luo, Q., Durvasula, K., & Lin, Y. H. (2016). Inconsistent Consonantal Effects on F0 in Cantonese and Mandarin. *Tonal Aspects of Languages 2016*, 52-55.

Maekawa, K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. In *ICSLP*.

Melvin S., Clopper C., Gender Variation in Creaky Voice and Fundamental Frequency. In: Proceedings of ICPHS, 2015

Nolan ,F. (1983). *The Phonetic Bases of Speaker Rrecognition*. Cambridge University Press, Cambridge

Paeschke, A., Kienast, M., & Sendlmeier, W. F. (1999, August). F0-contours in emotional speech. In *Proc. ICPhS* (Vol. 99, pp. 929-933).

Paul Boersma & David Weenink. (2013), Praat: doing phonetics by computer [Computer program].Version 5.3.51, retrieved 2 June 2013 from http://www.praat.org/

Pépiot, E. (2014). Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. *Speech Prosody 7* (pp. 305-309).

Rose, P. (2002). *Forensic speaker identification*. CRC Press.

Schwab, S. & Goldman, Jean-Philippe. Do speakers show different F0 when they speak in different languages? The case of English, French and German? Speech Prosody. 2016.

Talkin, D. (2015). REAPER: Robust Epoch And Pitch EstimatoR. Retrieved from https://github.com/google/REAPER

Traunmüller, H., & Eriksson, A. (1995). The perceptual evaluation of F 0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, *97*(3), 1905-1915.

Voigt, R., Jurafsky, D. & Sumner, M. (2016). Between- and Within-Speaker Effects of Bilingualism on F0 Variation. *INTERSPEECH*.

Wickham. H., ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, *26*(3), e111-e116.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception a. *The Journal of the Acoustical Society of America*, *136*(3), 1320-1333.